



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/86, 5/10, A61K 48/00, C12R 1/92	A2	(11) International Publication Number: WO 95/16048 (43) International Publication Date: 15 June 1995 (15.06.95)
(21) International Application Number: PCT/CA94/00678 (22) International Filing Date: 9 December 1994 (09.12.94) (30) Priority Data: 08/164,292 9 December 1993 (09.12.93) US (71) Applicant: UNIVERSITY OF SASKATCHEWAN [CA/CA]; 124 Veterinary Road, Saskatoon, Saskatchewan S5R 2Y4 (CA). (72) Inventors: MITTAL, Suresh, K.; #201-235 Kingsmere Boulevard, Saskatoon, Saskatchewan S7J 4J6 (CA). GRAHAM, Frank, L.; 34 Amelia Street, Hamilton, Ontario L8P 2V4 (CA). PREVEC, Ludvik; 944 LaSalle Park Road, Burlington, Ontario L7T 1M9 (CA). BABIUK, Lorne, A.; 245 East Place, Saskatoon, Saskatchewan S7J 2Y1 (CA). (74) Agent: VAN ZANT, Joan, M.; Scott & Ayles, 60 Queen Street, Ottawa, Ontario K1P 5Y7 (CA).		(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>Without international search report and to be republished upon receipt of that report.</i>
(54) Title: RECOMBINANT PROTEIN PRODUCTION IN BOVINE ADENOVIRUS EXPRESSION VECTOR SYSTEM (57) Abstract The present invention relates to novel live bovine adenovirus (BAV) expression vector systems in which part or all of one or both of the early region 1 (E1) and early region 3 (E3) genes are deleted and replaced by a foreign gene or fragment thereof and novel recombinant mammalian cell lines stably transformed with BAV E1 sequences, and therefore, express E1 gene products capable of allowing replication therein of a bovine adenovirus having an E1 deletion replaced by a heterologous nucleotide sequence encoding a foreign gene or fragment thereof and their use in production of (antigenic) polypeptides or fragments thereof for the purpose of live recombinant virus or subunit vaccine or for other therapies.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LV	Latvia	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

5 RECOMBINANT PROTEIN PRODUCTION IN BOVINE ADENOVIRUS
 EXPRESSION VECTOR SYSTEM

Technical Field

 The present invention relates novel bovine
10 adenovirus (BAV) expression vector systems in which
 one or both of the early region 1 (E1) and the early
 region 3 (E3) gene deletions are replaced by a foreign
 gene and novel recombinant mammalian cell lines stably
 transformed with BAV E1 sequences, and therefore,
15 expresses E1 gene products, to allow a bovine
 adenovirus with an E1 gene deletion replaced by a
 foreign gene to replicate therein. These materials
 are used in production of recombinant BAV expressing
 heterologous (antigenic) polypeptides or fragments for
20 the purpose of live recombinant virus or subunit
 vaccines or for other therapies.

Background of the Invention

 The adenoviruses cause enteric or
25 respiratory infection in humans as well as in domestic
 and laboratory animals.

 The bovine adenoviruses (BAVs) comprise at
 least nine serotypes divided into two subgroups.
 These subgroups have been characterized based on
30 enzyme-linked immunoassays (ELISA), serologic studies
 with immunofluorescence assays, virus-neutralization
 tests, immunoelectron microscopy, by their host
 specificity and clinical syndromes. Subgroup 1
 viruses include BAV 1, 2, 3 and 9 and grow relatively
35 well in established bovine cells compared to subgroup
 2 which includes BAV 4, 5, 6, 7 and 8.

 BAV3 was first isolated in 1965 and is the
 best characterized of the BAV genotypes and contains a

genome of approximately 35 kb (Kurokawa et al (1978) J. Virol. 28:212-218). The locations of hexon (Hu et al (1984) J. Virol. 49:604-608) and proteinase (Cai et al., (1990) Nuc. Acids Res., 18:5568), genes in the
5 BAV3 genome have been identified and sequenced. However, the location and sequences of other genes such as early region 1 (E1) and 3 (E3) in the BAV genome have not been reported.

In the human adenovirus (HAd) genome there
10 are two important regions: E1 and E3 in which foreign genes can be inserted to generate recombinant adenoviruses (Berkner and Sharp (1984) Nuc. Acid Res., 12:1925-1941 and Haj-Ahmad and Graham (1986) J. Virol., 57:267-274). E1 proteins are essential for
15 virus replication in tissue culture, however, conditional-helper adenovirus recombinants containing foreign DNA in the E1 region, can be generated in a cell line which constitutively expresses E1 (Graham et al., (1977) J. Gen. Virol., 36:59-72). In contrast,
20 E3 gene products of HAd 2 and HAd 5 are not required for in vitro or in vivo infectious virion production, but have an important role in host immune responses to virus infection (Andersson et al (1985) Cell 43:215-222; Burgert et al (1987) EMBO J. 6:2019-2026; Carlin et al (1989) Cell 57:135-144; Ginsberg et al (1989) PNAS, USA 86:3823-3827; Gooding et al (1988) Cell 53:341-346; Tollefson et al (1991) J. Virol. 65:3095-3105; Wold and Gooding (1989) Mol. Biol. Med. 6:433-452 and Wold and Gooding (1991) Virology 184:1-8).
25 The E3-19kiloDalton (kDa) glycoprotein (gp19) of human adenovirus type 2 (HAd2) binds to the heavy chain of a number of class 1 major histocompatibility complex (MHC) antigens in the endoplasmic reticulum thus inhibiting their transport to the plasma membrane
30 (Andersson et al. (1985) Cell 43:215-222; Burgert and Kvist, (1985) Cell 41:987-997; Burgert and Kvist, (1987) EMBO J. 6:2019-2026). The E3-14.7kDa protein of HAd2 or HAd5 prevents lysis of virus-infected mouse cells by tumor necrosis factor (TNF) (Gooding et al.

(1988) Cell 53:341-346). In addition, the E3-10.4kDa and E3-14.5kDa proteins form a complex to induce endosomal-mediated internalization and degradation of the epidermal growth factor receptor (EGF-R) in virus-
5 infected cells (Carlin et al. Cell 57:135-144; Tollefson et al. (1991) J. Virol. 65:3095-3105). The helper-independent recombinant adenoviruses having foreign genes in the E3 region replicate and express very well in every permissive cell line (Chanda et al
10 (1990) Virology 175:535-547; Dewar et al (1989) J. Virol. 63:129-136; Johnson et al (1988) Virology 164:1-14; Lubeck et al (1989) PNAS, USA 86:6763-6767; McDermott et al (1989) Virology 169:244-247; Mittal et al (1993) Virus Res. 28:67-90; Morin et al (1987)
15 PNAS, USA 84:4626-4630; Prevec et al (1990) J. Inf. Dis. 161:27-30; Prevec et al (1989) J. Gen. Virol. 70:429-434; Schneider et al (1989) J. Gen. Virol. 70:417-427 and Yuasa et al (1991) J. Gen. Virol. 72:1927-1934). Based on the above studies and the
20 suggestion that adenoviruses can package approximately 105% of the wild-type (wt) adenovirus genome (Bett et al (1993) J. Virol. 67:5911-5921 and Ghosh-Choudhury et al (1987) EMBO. J. 6:1733-1739), an insertion of up to 1.8 kb foreign DNA can be packaged into adenovirus
25 particles for use as an expression vector for foreign proteins without any compensating deletion.

It is assumed that an indigenous adenovirus vector would be better suited for use as a live recombinant virus vaccine in different animal species
30 compared to an adenovirus of human origin. Non-human adenovirus-based expression vectors have not been reported so far. If like HAd5 E3, the E3 regions in other adenoviruses are not essential for virus replication in cultured cells, adenovirus recombinants
35 containing foreign gene inserts in the E3 region could be generated.

BAV3 is a common pathogen of cattle usually resulting in subclinical infection though occasionally associated with a more serious respiratory tract

-4-

infection (Darbyshire et al., 1966 Res. Vet Sci 7:81-93; Mattson et al., 1988 J. Vet Res 49:67-69). BAV3 can produce tumors when injected into hamsters (Darbyshire, 1966 Nature 211:102) and viral DNA can efficiently effect morphological transformation of mouse, hamster or rat cells in culture (Tsukamoto and Sugino, 1972 J. Virol. 9:465-473; Motoi et al., 1972 Gann 63:415-418; M. Hitt, personal communication). Cross hybridization was observed between BAV3 and human adenovirus type 2 (HAd2) (Hu et al., 1984 J. Virol. 49:604-608) in most regions of the genome including some regions near but not at the left end of the genome.

The E1A gene products of the group C human adenoviruses have been very extensively studied and shown to mediate transactivation of both viral and cellular genes (Berk et al., 1979 Cell 17:935-944; Jones and Shenk, 1979 Cell 16:683-689; Nevins, 1981 Cell 26:213-220; Nevins, 1982 Cell 29:913-919; reviewed in Berk, 1986 Ann. Res. Genet 20:45-79), to effect transformation of cells in culture (reviewed in Graham, F.L. (1984) "Transformation by and oncogenicity of human adenoviruses. In: The Adenoviruses." H.S. Ginsberg, Editor. Plenum Press, New York; Branton et al., 1985 Biochim. Biophys. Acta 780:67-94) and induce cell DNA synthesis and mitosis (Zerler et al., 1987 Mol. Cell Biol. 7:821-929; Bellet et al., 1989 J. Virol. 63:303-310; Howe et al., 1990 PNAS, USA 87:5883-5887; Howe and Bayley, 1992 Virology 186:15-24). The E1A transcription unit comprises two coding sequences separated by an intron region which is deleted from all processed E1A transcripts. In the two largest mRNA species produced from the E1A transcription unit, the first coding regions is further subdivided into exon 1, a sequence found in both the 12s and 13s mRNA species, and the unique region, which is found only in the 13s mRNA species.

By

-5-

comparisons between E1A proteins of human and simian adenoviruses three regions of somewhat conserved protein sequence (CR) have been defined (Kimelman et al., 1985 J. Virol. 53:399-409). CR1 and CR2 are encoded in exon 1, while CR3 is encoded in the unique sequence and a small portion of exon 2. Binding sites for a number of cellular proteins including the retinoblastoma protein Rb, cyclin A and an associated protein kinase p33^{cdk2}, and other, as yet unassigned, proteins have been defined in exon 1 encoded regions of E1A proteins (Yee and Branton, 1985 Virology 147:142-153; Harlow et al., 1986 Mol. Cell Biol. 6:1579-1589; Barbeau et al., 1992 Biochem. Cell Biol. 70:1123-1134). Interaction of E1A with these cellular proteins has been implicated as the mechanism through which E1A participates in immortalization and oncogenic transformation (Egan et al, 1989 Oncogene 4:383-388; Whyte et al., 1988 Nature 334:124-129; Whyte et al, 1988 J. Virol. 62:257-265). While E1A alone may transform or immortalize cells in culture, the coexpression of both E1A and either the E1B-19k protein or the E1B-55k protein separately or together is usually required for high frequency transformation of rodent cells in culture (reviewed in Graham, 1984 *supra*; Branton et al., 1985 *supra*; McLorie et al., 1991 J. Gen Virol. 72:1467-1471).

Transactivation of other viral early genes in permissive infection of human cells is principally mediated by the amino acid sequence encoded in the CR3 region of E1A (Lillie et al., 1986 Cell 46:1043-1051). Conserved cysteine residues in a CysX₂CysX₁₃CysX₂Cys sequence motif in the unique region are associated with metal ion binding activity (Berg, 1986 *supra*) and are essential for transactivation activity (Jelsma et al., 1988 Virology 163:494-502; Culp et al., 1988 PNAS, USA 85:6450-6454). As well, the amino acids in CR3 which are immediately amino (N)-terminal to the metal binding domain have been shown to be important

-6-

in transcription activation, while those immediately carboxy (C)-terminal to the metal binding domain are important in forming associations with the promoter region (Lillie and Green, 1989 Nature 338:39-44; see Fig. 3).

The application of genetic engineering has resulted in several attempts to prepare adenovirus expression systems for obtaining vaccines. Examples of such research include the disclosures in U.S. patent 4,510,245 on an adenovirus major late promoter for expression in a yeast host; U.S. patent 4,920,209 on a live recombinant adenovirus type 7 with a gene coding for hepatitis-B surface antigen located at a deleted early region 3; European patent 389 286 on a non-defective human adenovirus 5 recombinant expression system in human cells for HCMV major envelope glycoprotein; WO 91/11525 on live non-pathogenic immunogenic viable canine adenovirus in a cell expressing E1a proteins; French patent 2 642 767 on vectors containing a leader and/or promoter from the E3 of adenovirus 2.

The selection of a suitable virus to act as a vector for foreign gene expression, and the identification of a suitable non-essential region as a site for insertion of the gene pose a challenge. In particular, the insertion site must be non-essential for the viable replication of the virus and its effective operation in tissue culture and also in vivo. Moreover, the insertion site must be capable of accepting new genetic material, whilst ensuring that the virus continues to replicate. An essential region of a virus genome can also be utilized for foreign gene insertion if the recombinant virus is grown in a cell line which complements the function of that particular essential region in trans.

The present inventors have now identified suitable regions in the BAV genome and have succeeded in inserting foreign genes to generate BAV recombinants.

Disclosure of the Invention

The present invention relates to novel bovine adenovirus expression vector systems in which part or all of one or both of the E1 and E3 gene regions are deleted and to recombinant mammalian cell lines of bovine origin transformed with the BAV E1 sequences, and thus, constitutively express the E1 gene products to allow bovine adenovirus, having a deletion of part or all of the E1 gene region replaced by a heterologous nucleotide sequence encoding a foreign gene or fragment thereof, to replicate therein and use of these materials in production of heterologous (antigenic) polypeptides or fragments thereof.

The invention also related to a method of preparing a live recombinant virus or subunit vaccines for producing antibodies or cell mediated immunity to an infectious organism in a mammal, such as bovine, which comprises inserting into the bovine adenovirus genome the gene or fragment coding for the antigen which corresponds to said antibodies or induces said cell mediated immunity, together with or without an effective promoter therefore, to produce BAV recombinants.

Generally, the foreign gene construct is cloned into a nucleotide sequence which represents only a part of the entire viral genome having one or more appropriate deletions. This chimeric DNA sequence is usually present in a plasmid which allows successful cloning to produce many copies of the sequence. The cloned foreign gene construct can then be included in the complete viral genome, for example, by in vivo recombination following a DNA-mediated cotransfection technique. Multiple copies of a coding sequence or more than one coding sequences can be inserted so that the recombinant vector can express more than one foreign protein. The foreign gene can have additions, deletions or substitutions to enhance

-8-

expression and/or immunological effects of the expressed protein.

The invention also includes an expression system comprising an bovine adenovirus expression
5 vector wherein heterologous nucleotide sequences with or without any exogenous regulatory elements, replace the E1 gene region and/or part or all of the E3 gene region.

The invention also includes (A) a
10 recombinant vector system comprising the entire BAV DNA and a plasmid or two plasmids capable of generating a recombinant virus by in vivo recombination following cotransfection of a suitable cell line comprising BAV DNA representing the entire
15 wild-type BAV genome and a plasmid comprising a bovine adenovirus left or right end sequences containing the E1 or E3 gene regions, respectively, with a heterologous nucleotide sequence encoding a foreign gene or fragment thereof substituted for part or all
20 of the E1 or E3 gene regions; (B) a live recombinant bovine adenovirus vector (BAV) system selected from the group consisting of: (a) a system wherein part or all of the E1 gene region is replaced by a heterologous nucleotide sequence encoding a foreign
25 gene or fragment thereof; (b) a system wherein a part or all of the E3 gene region is replaced by a heterologous nucleotide sequence encoding a foreign gene or fragment thereof; and (c) a system wherein part or all of the E1 gene region and part or all of
30 the E3 gene region are deleted and a heterologous nucleotide sequence encoding a foreign gene or fragment thereof is inserted into at least one of the deletions; (C) a recombinant bovine adenovirus (BAV) comprising a deletion of part or all of E1 gene
35 region, a deletion of part or all of E3 gene region or deletion of both, and inserted into at least one deletion a heterologous nucleotide sequence coding for an antigenic determinant of a disease causing organism; (D) a recombinant bovine adenovirus

expression system comprising a deletion of part or all of E1, a deletion of part or all of E3, or both deletions, and inserted into at least one deletion a heterologous nucleotide sequence coding for a foreign gene or fragment thereof under control of an expression promoter: or (E) a recombinant bovine adenovirus (BAV) for producing an immune response in a mammalian host comprising: (1) BAV recombinant containing a heterologous nucleotide sequence coding for an antigenic determinant needed to obtain the desired immune response in association with or without (2) an effective promoter to provide expression of said antigenic determinant in immunogenic quantities for use as a live recombinant virus or recombinant protein or subunit vaccine; (F) a mutant bovine adenovirus (BAV) comprising a deletion of part or all of E1 and/or a deletion of part or all of E3.

Recombinant mammalian cell lines stably transformed with BAV E1 gene region sequences, said recombinant cell lines thereby capable of allowing replication therein of a bovine adenovirus comprising a deletion of part or all of the E1 or E3 gene regions replaced by a heterologous or homologous nucleotide sequence encoding a foreign gene or fragment thereof. The invention also includes production, isolation and purification of polypeptides or fragments thereof, such as growth factors, receptors and other cellular proteins from recombinant bovine cell lines expressing BAV E1 gene products.

The invention also includes a method for providing gene therapy to a mammal in need thereof to control a gene deficiency which comprises administering to said mammal a live recombinant bovine adenovirus containing a foreign nucleotide sequence encoding a non-defective form of said gene under conditions wherein the recombinant virus vector genome is incorporated into said mammalian genome or is maintained independently and extrachromosomally to

-10-

provide expression of the required gene in a target organ or tissue.

Another aspect of the invention provides a virus vaccine composition which comprises the recombinant virus or recombinant protein in association with or without a pharmaceutically acceptable carrier. The recombinant virus vaccine can be formulated for administration by an oral dosage (e.g. as an enteric coated tablet), by injection or otherwise. More specifically, these include a vaccine for protecting a mammalian host against infection comprising a live recombinant adenovirus or recombinant protein produced by the recombinant adenovirus of the invention wherein the foreign gene or fragment encodes an antigen and formulated with or without a pharmaceutically acceptable carrier.

The invention also includes methods of producing antibodies or cell mediated immunity in a mammal including (1) a method for eliciting an immune response in a mammalian host against an infection comprising: administering a vaccine comprising a live BAV recombinant of the invention wherein the foreign gene or fragment encodes an antigen with or without a pharmaceutically acceptable carrier, and (2) a method for eliciting an immune response in a mammalian host against an infection comprising: administering a vaccine comprising a recombinant antigen prepared by culturing a BAV recombinant wherein the foreign gene or fragment encodes the desired antigen with or without a pharmaceutically acceptable carrier.

The following disclosure will render these and other embodiments of the present invention readily apparent to those of skill in the art. While the disclosure often refers to bovine adenovirus type 3 (BAV3), it should be understood that this is for the purpose of illustration and that the same features apply to bovine adenovirus of the other type, 1, 2, 4, 5, 6, 7 8, and 9 and the invention described and

-11-

claimed herein is intended to cover all of these bovine adenovirus types.

Brief Description of the Drawings

5 Figure 1. Sequence and major open reading frames of the left 11% of the BAV3 genome. The region comprises the E1 and protein IX transcription region. The 195 nucleotide inverted terminal repeat sequence identified by Shinagawa et al., 1987 Gene 55:85-93 is
10 shown in *italics*. The amino acid sequence for the largest E1A protein, two E1B proteins and protein IX are presented. The probable splice donor ([), splice acceptor (]) and intron sequence (*underlined italics*) within the E1A region are marked. A 35 base pair
15 repeat sequence between E1A and E1B is indicated in **bold underline**. Possible transcription promoter TATA sequences and possible poly A addition sequences AATAA are also indicated.

Figure 2. Regions of homology in the E1A
20 proteins of BAV3 and human adenovirus type 5 (HAd5). The amino acid residue of each serotype is indicated. A. Conserved region 3 (CR3) of HAd5 subdivided into three functional regions as defined by Lillie et al (1989) Nature 338:39-44 and described in the
25 Background of the Invention. The intron sequence of BAV3 E1A occurs within the serine amino acid codon at position 204. B. A portion of conserved region 2 (CR2) of HAd5, showing the residues thought to be important in the binding of retinoblastoma protein Rb
30 (Dyson et al., 1990 J. Virol. 64:1353-1356), and the comparable sequence from BAV3.

Figure 3. Homology regions between the HAd5 and E1B 19k (176R) protein and the corresponding BAV3 (157R) protein. The amino acid residue number for
35 each of the viruses is indicated.

Figure 4. The C-terminal 346R of HAd5 E1B 56k (496R) and the corresponding BAV3 protein (420R). The HAd5 protein comparison begins at residue 150 and the BAV3 (in *italics*) at residue 74.

-12-

The amino terminal regions of these proteins which are not presented show no significant homology.

Figure 5. Homology comparison of the amino acid sequence of HAd5 protein IX and the corresponding protein of BAV3 (in italics).

Figure 6. The genome of BAV3 showing the location of *EcoRI*, *XbaI* and *BAMHI* sites and the structure of the 5100bp segment from 77 to 92 m.u. ORFs for the upper strand which can encode 60 amino acids or more are represented by bars. Shaded portions indicate regions of similarity to pVIII, 14.7K E3 and fibre proteins of HAd2 or -5. The first methionine followed by a stretch of amino acids of at least 50 is shown by an open triangle. Termination codons for ORFs likely to code for viral proteins are shown by closed triangles.

Figure 7. Nucleotide sequence of BAV3 between 77 and 92 m.u. showing ORFs that have the potential to encode polypeptides of at least 50 amino acids after the initiating methionine. The nucleotide sequence was analyzed using the program DISPCOD (PC/GENE). Potential N-glycosylation sites (N-X-T/S) and polyadenylation signals are underlined and the first methionine of each ORF is shown in bold.

Figure 8. Comparison between the predicted amino acid sequences for the ORFs of BAV3 and known proteins of HAd2 or -5 using the computer program PALIGN (PC/GENE), with comparison matrix structural-genetic matrix; open gap cost 6; unit gap cost 2. Identical residues are indicated by a colon and similar residues by a dot. (a) Comparison between the predicted amino acid sequence encoded by the 3' end of BAV3 ORF 1 and the HAd2 hexon-associated pVIII precursor. (b) Comparison between the ORF 4 and the HAd5 14.7K E3 protein. (c) Comparison between the predicted amino acid sequence encoded by BAV3 ORF 6 and the HAd2 fibre protein.

Figure 9. Construction of BAV3 E3 transfer vector containing the firefly luciferase gene. The

-13-

3.0 kb BamHI 'D' fragment of the BAV3 genome which falls between m.u. 77.8 and 86.4, contains almost the entire E3 region (Mittal et al (1992) J. Gen. Virol. 73:3295-3000). This 3.0 kb fragment was isolated by
5 digesting BAV3 DNA with BamHI and cloned into pUC18 at the BamHI site to obtain pSM14. Similarly, the 4.8 kb BamHI 'C' fragment of BAV3 DNA which extends between m.u. 86.4 and 100 was isolated and inserted into pUC18 to produce pSM17. To delete a 696 bp XhoI-NcoI
10 fragment, pSM14 was cleaved with XhoI and NcoI, the larger fragment was purified and the ends were made blunt with Klenow fragment of DNA polymerase I and a NruI-SalI linker was inserted to generate pSM14del12. A 2.3 kb BamHI fragment
15 containing BAV3 sequences, an E3 deletion and NruI and SalI cloning sites, was inserted into pSM17 at the BamHI site to obtain pSM41, however, this step was not required for construction of a BAV3 E3 transfer vector. A 1716 bp fragment containing the firefly
20 luciferase gene (de Wet et al (1987) Mol. Cell. Biol. 7:725-737) was isolated by digesting pSVOA/L (provided by D. R. Helinski, University of California at San Diego, CA) with BsmI and SspI as described (Mittal et al (1993) Virus Res. 28:67-90), and the ends were made
25 blunt with Klenow. The luciferase gene was inserted into pSM41 at the SalI site by blunt end ligation. The resultant plasmid was named pSM41-Luc which contained the luciferase gene in the same orientation as the E3 transcription unit. The plasmid pKN30 was
30 digested with XbaI and inserted into pSM41-Luc (partially cleaved with XbaI) at a XbaI site present within the luciferase gene to obtain pSM41-Luc-Kan. The plasmid pSM14 was digested with BamHI and a 3.0 kb fragment was isolated and inserted into pSM17 at the
35 BamHI site to generate pSM43. The 18.5 kb XbaI 'A' fragment of the BAV3 genome which falls between m.u. 31.5 and 84.3 was cloned into pUC18 at the XbaI site to result pSM21. A 18.5 kb XbaI fragment was purified from pSM21 after cleavage with XbaI and

-14-

inserted into pSM43 at the XbaI site and the resultant plasmid was named pSM51. A

7.7 kb BamHI fragment containing the luciferase gene and kan^r gene was isolated after digesting pSM41-Luc-Kan with BamHI and ligated to pSM51, partially digested with BamHI, to isolate pSM51-Luc-Kan in the presence of ampicillin and kanamycin. Finally the kan^r gene was deleted from pSM51-Luc-Kan by partial cleavage with XbaI and religation to obtain pSM51-Luc.

Figure 10. Generation of BAV3 recombinants containing the firefly luciferase in the E3 region. The plasmid pSM51-Luc contains the BAV3 genome between m.u. 77.8-84.3 and 31.5-100, a 696 bp deletion in E3 and the luciferase gene in E3 in the E3 parallel orientation. The BAV3 genome digested with PvuI and uncut pSM51-Luc were used for cotransfection of MDBK cells transformed with a plasmid containing BAV3 E1 sequences to rescue the luciferase gene in E3 of the BAV3 genome by *in vivo* recombination. The resulting BAV3-luciferase recombinants (BAV3-Luc) isolated from two independent experiments were named BAV3-Luc (3.1) and BAV3-Luc (3.2). The BamHI restriction map of the BAV3-Luc genome is shown. The position and orientation of the firefly luciferase gene is shown as a hatched arrow.

Figure 11. Southern blot analyses of restriction enzymes digested DNA fragments of the wt BAV3 or recombinant genomes by using a 696 bp XhoI-NcoI fragment from pSM14 (Fig. 9) and a DNA fragment containing the luciferase gene as probes. 100 ng DNA isolated from the mock (lanes 1, 2, 3), BAV3-Luc (3.1) (lanes 4, 5, 6), BAV3-Luc (3.2) (lanes 7, 8, 9) or wt BAV3 (lanes 10, 11, 12)-infected MDBK cells were digested with BamHI (lanes 1, 4, 7, 10), EcoRI (lanes 2, 5, 8, 11) or XbaI (lanes 3, 6, 9, 12) and analyzed by agarose gel electrophoresis. The DNA fragments from the gel were transferred onto a GeneScreenPlus™ membrane and hybridized with a 696 bp XhoI-NcoI

-15-

fragment from pSM14 (Fig. 9) labeled with ^{32}P using Pharmacia Oligolabeling Kit (panel A). Panel B blot represents duplicate samples as in panel A but was probed with a 1716 bp BsmI-SspI fragment containing the luciferase gene (Fig. 9). The sizes of bands visualized following hybridization are shown in kb on the right in panel A and on the left in panel B.

B: BamHI, E: EcoRI, Xb: XbaI, 3.1: BAV3-Luc (3.1), 3.2: BAV3-Luc (3.2) and wt: wild-type BAV3.

Figure 12. Single step growth curve for wt BAV3 and BAV3-Luc. Confluent monolayers of MDBK cells in 25 mm multi-well culture plates were inoculated with the wt BAV3, BAV3-Luc (3.1) or BAV3-Luc (3.2) at a m.o.i. of 10 p.f.u. per cell. The virus was allowed to adsorb for 1 h at 37°C, cell monolayers were washed 3 times with PBS⁺⁺ (0.137 M NaCl, 2.7 mM KCl, 8 mM Na₂HPO₄, 1.5 mM KH₂PO₄, containing 0.01% CaCl₂·2H₂O & 0.01% MgCl₂·6H₂O) and incubated at 37°C in 1 ml maintenance medium containing 2% horse serum. At various times post-infection, cells were harvested along with the supernatant, frozen and thawed three times and titrated on MDBK cells by plaque assay. Results are the means of duplicate samples.

Figure 13. Kinetics of luciferase expression in MDBK cells-infected with BAV3-Luc. Confluent MDBK cell monolayers in 25 mm multi-well culture plates were infected with BAV3-Luc (3.1) or BAV3-Luc (3.2) at a m.o.i. of 50 p.f.u. per cell. At indicated time points post-infection, virus-infected cells were harvested and assayed in duplicate for luciferase activity.

Figure 14. Luciferase expression in the presence of 1- β -D-arabinofluranosyl cytosine (AraC) in MDBK cells-infected with BAV3-Luc. Confluent MDBK cell monolayers in 25 mm multi-well culture plates were infected with A) BAV3-Luc (3.1) or B) BAV3-Luc (3.2) at a m.o.i. of 50 p.f.u. per cell and incubated in the absence or presence of 50 μg AraC per ml of

-16-

maintenance medium. At indicated time points post-infection, virus-infected cells were harvested and assayed in duplicate for luciferase activity.

Figure 15. Transcription maps of the wt BAV3 and BAV3-Luc genomes in the E3 region. The genome of wt BAV3 between m.u. 77 and 82 is shown which represents the E3 region. The location of XhoI and NcoI sites which were used to make an E3 deletion are shown. (a) The three frames (F1, F2 and F3) representing the open reading frames (ORFs) in the upper strand of the wt BAV3 genome in the E3 region are represented by bars. The shaded portions indicate regions of similarities to pVIII and E3-14.7 kDa proteins of HAd5. The positions of the initiation and termination codons for ORFs likely to code for viral proteins are shown by open and closed triangles, respectively. (b) The predicted ORFs for the upper strand in E3 of the BAV3-Luc genome are shown after a 696 bp XhoI-NcoI E3 deletion replaced by the luciferase gene. The ORFs for pVIII and E3-14.7 kDa proteins are intact. The transcription map of the wt BAV3 E3 was adapted from the DNA sequence submitted to the GenBank database under accession number D16839.

Figure 16. Western blot analysis of virus-infected MDBK cells using an anti-luciferase antibody. Confluent monolayers of MDBK cells were mock-infected (lane 1) or infected with the wt BAV3 (lane 2), BAV3-Luc (3.1) (lane 3) and BAV3-Luc (3.2) (lane 4) at a m.o.i. of 50 p.f.u. per cell, harvested at 18 h post-infection, cell extracts prepared and analyzed by SDS-PAGE and Western blotting using a rabbit anti-luciferase antibody. Purified firefly luciferase was used as a positive control (lane 5). The lane 5 was excised to obtain a shorter exposure. The protein molecular weight markers in kDa are shown on the left. The arrow indicates the 62 kDa luciferase bands reacted with the anti-luciferase antibody. wt: wild-type BAV3, 3.1: BAV3-Luc (3.1) and 3.2: BAV3-Luc (3.2).

-17-

Figure 17. Construction of pSM71-neo. A 8.4 kb SalI fragment of the BAV3 genome which falls between m.u. 0 and 24 was isolated and inserted into pUC19 at the SalI-SmaI site to generate pSM71. The
5 plasmid, pRSDneo (Fitzpatrick et al (1990) Virology 176:145-157) contains the neomycin-resistant (neo^r gene flanked with the simian virus 40 (SV40) regulatory sequences originally from the plasmid, pSV2neo (Southern et al (1982) J. Mol. Appl. Genet 1:327-341)
10 after deleting a portion of the SV40 sequences upstream of the neo^r gene to remove several false initiation codons. A 2.6 kb fragment containing the neo^r gene under the control of the SV40 regulatory sequences, was obtained from the plasmid, pRSDneo
15 after digestion with BamHI and BglII, and cloned into pSM71 at the SalI site by blunt end ligation to obtain pSM71-neo containing the neo^r gene in the E1 parallel orientation.

Figure 18. Construction of pSM61-kan 1 and
20 pSM61-kan2. A 11.9 kb BglII fragment of the BAV3 genome which extends between m.u. 0 and 34 was purified and introduced into pUC19 at the BamHI-HincII site to obtain pSM61. The plasmid, pKN30 contains the neo^r gene along with SV40 promoter and polyadenylation
25 sequences from the plasmid pSV2neo without any modification. The entire pKN30 plasmid was inserted into pSM61 at the SalI site to generate pSM61-kan1 having the neo^r gene in the E1 anti-parallel orientation and pSM61-kan2 when the neo^r gene is in the
30 E1 parallel orientation.

Figure 19. Construction of an E1 transfer plasmid containing the beta-galactosidase gene.

The plasmid, pSM71 which contains the BAV3 genome between m.u. 0 and 24, was cleaved with ClaI
35 and partially with AvrII to delete a 2.6 kb AvrII-ClaI fragment (between m.u. 1.3 and 8.7) which falls within the E1 region. A 0.5 kb fragment containing the SV40 promoter and polyadenylation sequences was obtained

from pFG144K5-SV by digesting with XbaI and inserted into pSM71 to replace the 2.6 kb deletion to generate pSM71-dell-SV. A 3.26 kb fragment containing the bacterial beta-galactosidase gene was isolated from pDUC/Z (Liang et al (1993) Virology 195:42-50) after cleavage with NcoI and HindIII and cloned into pSM71-dell-SV at the BamHI site to put the beta-galactosidase gene under the control of the SV40 regulatory sequences to obtain pSM71-Z.

10

Modes of Carrying Out the Invention

The practice of the present invention will employ, unless otherwise indicated, conventional microbiology, immunology, virology, molecular biology, and recombinant DNA techniques which are within the skill of the art. These techniques are fully explained in the literature. See, e.g., Maniatis et al., Molecular Cloning: A Laboratory Manual (1982); DNA Cloning: A Practical Approach, vols. I & II (D. Glover, ed.); Oligonucleotide Synthesis (N. Gait, ed. (1984)); Nucleic Acid Hybridization (B. Hames & S. Higgins, eds. (1985)); Transcription and Translation (B. Hames & S. Higgins, eds. (1984)); Animal Cell Culture (R. Freshney, ed. (1986)); Perbal, A Practical Guide to Molecular Cloning (1984). Sambrook et al., Molecular Cloning: A Laboratory Manual (2nd Edition); vols. I, II & III (1989).

20
25
30

A. Definitions

In describing the present invention, the following terminology, as defined below, will be used.

A "replicon" is any genetic element (e.g., plasmid, chromosome, virus) that functions as an autonomous unit of DNA replication in vivo; i.e., is capable of replication under its own control.

A "vector" is a replicon, such as a plasmid, phage, cosmid or virus, to which another DNA segment

-19-

may be attached so as to bring about the replication of the attached segment.

By "live virus" is meant, in contradistinction to "killed" virus, a virus which is capable of producing identical progeny in tissue culture and inoculated animals.

A "helper-free virus vector" is a vector that does not require a second virus or a cell line to supply something defective in the vector.

A "double-stranded DNA molecule" refers to the polymeric form of deoxyribonucleotides (adenine, guanine, thymine, or cytosine) in its normal, double-stranded helix. This term refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, inter alia, in linear DNA molecules (e.g., restriction fragments of DNA from viruses, plasmids, and chromosomes). In discussing the structure of particular double-stranded DNA molecules, sequences may be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the nontranscribed strand of DNA (i.e., the strand having the sequence homologous to the mRNA).

A DNA "coding sequence" is a DNA sequence which is transcribed and translated into a polypeptide in vivo when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but is not limited to, procaryotic sequences, cDNA from eucaryotic mRNA, genomic DNA sequences from eucaryotic (e.g., mammalian) DNA, viral DNA, and even synthetic DNA sequences. A polyadenylation signal and transcription termination sequence will usually be located 3' to the coding sequence.

A "transcriptional promoter sequence" is a DNA regulatory region capable of binding RNA

-20-

polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of defining the present invention, the promoter sequence is bound at the 3' terminus by the translation start codon (ATG) of a coding sequence and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site (conveniently defined by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eucaryotic promoters will often, but not always, contain "TATA" boxes and "CAAT" boxes. Procaryotic promoters contain Shine-Dalgarno sequences in addition to the -10 and -35 consensus sequences.

DNA "control sequences" refer collectively to promoter sequences, ribosome binding sites, polyadenylation signals, transcription termination sequences, upstream regulatory domains, enhancers, and the like, which collectively provide for the transcription and translation of a coding sequence in a host cell.

A coding sequence or sequence encoding is "operably linked to" or "under the control of" control sequences in a cell when RNA polymerase will bind the promoter sequence and transcribe the coding sequence into mRNA, which is then translated into the polypeptide encoded by the coding sequence.

A "host cell" is a cell which has been transformed, or is capable of transformation, by an exogenous DNA sequence.

A cell has been "transformed" by exogenous DNA when such exogenous DNA has been introduced inside the cell membrane. Exogenous DNA may or may not be integrated (covalently linked) to chromosomal DNA making up the genome of the cell. In procaryotes and yeasts, for example, the exogenous DNA may be

-21-

maintained on an episomal element, such as a plasmid. A stably transformed cell is one in which the exogenous DNA has become integrated into the chromosome so that it is inherited by daughter cells through chromosome replication. For mammalian cells, this stability is demonstrated by the ability of the cell to establish cell lines or clones comprised of a population of daughter cell containing the exogenous DNA.

10 A "clone" is a population of daughter cells derived from a single cell or common ancestor. A "cell line" is a clone of a primary cell that is capable of stable growth in vitro for many generations.

15 Two polypeptide sequences are "substantially homologous" when at least about 80% (preferably at least about 90%, and most preferably at least about 95%) of the amino acids match over a defined length of the molecule.

20 Two DNA sequences are "substantially homologous" when they are identical to or not differing in more than 40% of the nucleotides, more preferably about 20% of the nucleotides, and most preferably about 10% of the nucleotides.

25 DNA sequences that are substantially homologous can be identified in a Southern hybridization experiment under, for example, stringent conditions, as defined for that particular system. Defining appropriate hybridization conditions is within the skill of the art. See, e.g., Maniatis et al., supra; DNA Cloning, vols. I & II, supra; Nucleic Acid Hybridization, supra.

30 A "heterologous" region of a DNA construct is an identifiable segment of DNA within or attached to another DNA molecule that is not found in association with the other molecule in nature. Thus, when the heterologous region encodes a viral gene, the gene will usually be flanked by DNA that does not flank the viral gene in the genome of the source virus

-22-

or virus-infected cells. Another example of the heterologous coding sequence is a construct where the coding sequence itself is not found in nature (e.g., synthetic sequences having codons different from the native gene). Allelic variation or naturally occurring mutational events do not give rise to a heterologous region of DNA, as used herein.

"Bovine host" refers to cattle of any breed, adult or infant.

The term "protein" is used herein to designate a polypeptide or glycosylated polypeptide, respectively, unless otherwise noted. The term "polypeptide" is used in its broadest sense, i.e., any polymer of amino acids (dipeptide or greater) linked through peptide bonds. Thus, the term "polypeptide" includes proteins, oligopeptides, protein fragments, analogs, muteins, fusion proteins and the like.

"Fusion protein" is usually defined as the expression product of a gene comprising a first region encoding a leader sequence or a stabilizing polypeptide, and a second region encoding a heterologous protein. It involves a polypeptide comprising an antigenic protein fragment or a full length BAV protein sequence as well as (a) heterologous sequence(s), typically a leader sequence functional for secretion in a recombinant host for intracellularly expressed polypeptide, or an N-terminal sequence that protects the protein from host cell proteases, such as SOD. An antigenic protein fragment is usually about 5-7 amino acids in length.

"Native" proteins or polypeptides refer to proteins or polypeptides recovered from BAV or BAV-infected cells. Thus, the term "native BAV polypeptide" would include naturally occurring BAV proteins and fragments thereof. "Non-native" polypeptides refer to polypeptides that have been produced by recombinant DNA methods or by direct synthesis. "Recombinant" polypeptides refers to polypeptides produced by recombinant DNA techniques;

-23-

i.e., produced from cells transformed by an exogenous DNA construct encoding the desired polypeptide.

A "substantially pure" protein will be free of other proteins, preferably at least 10% homogeneous, more preferably 60% homogeneous, and most preferably 95% homogeneous.

An "antigen" refers to a molecule containing one or more epitopes that will stimulate a host's immune system to make a humoral and/or cellular antigen-specific response. The term is also used interchangeably with "immunogen."

A "hapten" is a molecule containing one or more epitopes that does not stimulate a host's immune system to make a humoral or cellular response unless linked to a carrier.

The term "epitope" refers to the site on an antigen or hapten to which a specific antibody molecule binds or is recognized by T cells. The term is also used interchangeably with "antigenic determinant" or "antigenic determinant site."

An "immunological response" to a composition or vaccine is the development in the host of a cellular and/or antibody-mediated immune response to the composition or vaccine of interest. Usually, such a response consists of the subject producing antibodies, B cells, helper T cells, suppressor T cells, and/or cytotoxic T cells directed specifically to an antigen or antigens included in the composition or vaccine of interest.

The terms "immunogenic polypeptide" and "immunogenic amino acid sequence" refer to a polypeptide or amino acid sequence, respectively, which elicit antibodies that neutralize viral infectivity, and/or mediate antibody-complement or antibody dependent cell cytotoxicity to provide protection of an immunized host. An "immunogenic polypeptide" as used herein, includes the full length (or near full length) sequence of the desired protein or an immunogenic fragment thereof.

-24-

By "immunogenic fragment" is meant a fragment of a polypeptide which includes one or more epitopes and thus elicits antibodies that neutralize viral infectivity, and/or mediates antibody-complement or antibody dependent cell cytotoxicity to provide protection of an immunized host. Such fragments will usually be at least about 5 amino acids in length, and preferably at least about 10 to 15 amino acids in length. There is no critical upper limit to the length of the fragment, which could comprise nearly the full length of the protein sequence, or even a fusion protein comprising fragments of two or more of the antigens. The term "treatment" as used herein refers to treatment of a mammal, such as bovine or the like, either (i) the prevention of infection or reinfection (prophylaxis), or (ii) the reduction or elimination of symptoms of an infection. The vaccine comprises the recombinant BAV itself or recombinant antigen produced by recombinant BAV.

By "infectious" is meant having the capacity to deliver the viral genome into cells.

B. General Method

The present invention identifies and provides a means of deleting part or all of the nucleotide sequence of bovine adenovirus E1 and/or E3 gene regions to provide sites into which heterologous or homologous nucleotide sequences encoding foreign genes or fragments thereof can be inserted to generate bovine adenovirus recombinants. By "deleting part of" the nucleotide sequence is meant using conventional genetic engineering techniques for deleting the nucleotide sequence of part of the E1 and/or E3 region.

Various foreign genes or coding sequences (prokaryotic, and eukaryotic) can be inserted in the bovine adenovirus nucleotide sequence, e.g., DNA, in accordance with the present invention, particularly to provide protection against a wide range of diseases

-25-

and many such genes are already known in the art. The problem heretofore having been to provide a safe, convenient and effective vaccine vector for the genes or coding sequences.

5 It is also possible that only fragments of nucleotide sequences of genes can be used (where these are sufficient to generate a protective immune response) rather than the complete sequence as found in the wild-type organism. Where available, synthetic
10 genes or fragments thereof can also be used. However, the present invention can be used with a wide variety of genes, fragment and the like, and is not limited to those set out above.

 In some cases the gene for a particular
15 antigen can contain a large number of introns or can be from an RNA virus, in these cases a complementary DNA copy (cDNA) can be used.

 In order for successful expression of the gene to occur, it can be inserted into an expression
20 vector together with a suitable promoter including enhancer elements and polyadenylation sequences. A number of eucaryotic promoter and polyadenylation sequences which provide successful expression of foreign genes in mammalian cells and how to construct
25 expression cassettes, are known in the art, for example in U.S. patent 5,151,267, the disclosures of which are incorporated herein by reference. The promoter is selected to give optimal expression of immunogenic protein which in turn satisfactorily leads
30 to humoral, cell mediated and mucosal immune responses according to known criteria.

 The foreign protein produced by expression *in vivo* in a recombinant virus-infected cell may be itself immunogenic. More than one foreign gene can be
35 inserted into the viral genome to obtain successful production of more than one effective protein.

 Thus with the recombinant virus of the present invention, it is possible to provide protection against a wide variety of diseases

-26-

affecting cattle. Any of the recombinant antigenic determinant or recombinant live virus of the invention can be formulated and used in substantially the same manner as described for the antigenic determinant
5 vaccines or an live vaccine vectors.

The antigens used in the present invention can be either native or recombinant antigenic polypeptides or fragments. They can be partial sequences, full-length sequences, or even fusions
10 (e.g., having appropriate leader sequences for the recombinant host, or with an additional antigen sequence for another pathogen). The preferred antigenic polypeptide to be expressed by the virus systems of the present invention contain full-length
15 (or near full-length) sequences encoding antigens. Alternatively, shorter sequences that are antigenic (i.e., encode one or more epitopes) can be used. The shorter sequence can encode a "neutralizing epitope," which is defined as an epitope capable of eliciting
20 antibodies that neutralize virus infectivity in an in vitro assay. Preferably the peptide should encode a "protective epitope" that is capable of raising in the host an "protective immune response;" i.e., an antibody- and/or a cell-mediated immune response that
25 protects an immunized host from infection.

The antigens used in the present invention, particularly when comprised of short oligopeptides, can be conjugated to a vaccine carrier. Vaccine carriers are well known in the art: for example,
30 bovine serum albumin (BSA), human serum albumin (HSA) and keyhole limpet hemocyanin (KLH). A preferred carrier protein, rotavirus VP6, is disclosed in EPO Pub. No. 0259149, the disclosure of which is incorporated by reference herein.

35 Genes for desired antigens or coding sequences thereof which can be inserted include those of organisms which cause disease in mammals, particularly bovine pathogens such as bovine rotavirus, bovine coronavirus, bovine herpes virus

-27-

type 1, bovine respiratory syncytial virus, bovine parainfluenza virus type 3 (BPI-3), bovine diarrhea virus, Pasteurella haemolytica, Haemophilus somnus and the like. The vaccines of the invention carrying
5 foreign genes or fragments can also be orally administered in a suitable oral carrier, such as in an enteric-coated dosage form. Oral formulations include such normally-employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch,
10 magnesium stearate, sodium saccharin cellulose, magnesium carbonate, and the like. Oral vaccine compositions may be taken in the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations, or powders, containing from
15 about 10% to about 95% of the active ingredient, preferably about 25% to about 70%. An oral vaccine may be preferable to raise mucosal immunity in combination with systemic immunity, which plays an important role in protection against pathogens
20 infecting the gastrointestinal tract.

In addition, the vaccine be formulated into a suppository. For suppositories, the vaccine composition will include traditional binders and carriers, such as polyalkaline glycols or
25 triglycerides. Such suppositories may be formed from mixtures containing the active ingredient in the range of about 0.5% to about 10% (w/w), preferably about 1% to about 2%.

Protocols for administering to animals the
30 vaccine composition(s) of the present invention are within the skill of the art in view of the present disclosure. Those skilled in the art will select a concentration of the vaccine composition in a dose effective to elicit an antibody and/or T-cell mediated
35 immune response to the antigenic fragment. Within wide limits, the dosage is not believed to be critical. Typically, the vaccine composition is administered in a manner which will deliver between about 1 to about 1,000 micrograms of the subunit

-28-

antigen in a convenient volume of vehicle, e.g., about 1-10 cc. Preferably, the dosage in a single immunization will deliver from about 1 to about 500 micrograms of subunit antigen, more preferably about 5-10 to about 100-200 micrograms (e.g., 5-200 micrograms).

The timing of administration may also be important. For example, a primary inoculation preferably may be followed by subsequent booster inoculations if needed. It may also be preferred, although optional, to administer a second, booster immunization to the animal several weeks to several months after the initial immunization. To insure sustained high levels of protection against disease, it may be helpful to readminister a booster immunization to the animals at regular intervals, for example once every several years. Alternatively, an initial dose may be administered orally followed by later inoculations, or vice versa. Preferred vaccination protocols can be established through routine vaccination protocol experiments.

The dosage for all routes of administration of *in vivo* recombinant virus vaccine depends on various factors including, the size of patient, nature of infection against which protection is needed, carrier and the like and can readily be determined by those of skill in the art. By way of non-limiting example, a dosage of between 10^3 pfu and 10^8 pfu and the like can be used. As with *in vitro* subunit vaccines, additional dosages can be given as determined by the clinical factors involved.

In one embodiment of the invention, a number of recombinant cell lines are produced according to the present invention by constructing an expression cassette comprising the BAV E1 region and transforming host cells therewith to provide cell lines or cultures expressing the E1 proteins. These recombinant cell lines are capable of allowing a recombinant BAV, having an E1 gene region deletion replaced by

-29-

heterologous nucleotide sequence encoding for a foreign gene or fragment, to replicate and express the desired foreign gene or fragment thereof which is encoded within the recombinant BAV. These cell lines
5 are also extremely useful in generating recombinant BAV, having an E3 gene deletion replaced by heterologous nucleotide sequence encoding for a foreign gene or fragment, by in vivo recombination following DNA-mediated cotransfection.

10 In one embodiment of the invention, the recombinant expression cassette can be obtained by cleaving the wild-type BAV genome with an appropriate restriction enzyme to produce a DNA fragment representing the left end or the right end of the
15 genome comprising E1 or E3 gene region sequences, respectively and inserting the left or right end fragment into a cloning vehicle, such as plasmid and thereafter inserting at least one DNA sequence encoding a foreign protein, into E1 or E3 deletion
20 with or without the control of an exogenous promoter. The recombinant expression cassette is contacted with the wild-type BAV DNA through homologous recombination or other conventional genetic engineering method within an E1 transformed cell line to obtain the
25 desired recombinant.

The invention also includes an expression system comprising an bovine adenovirus expression vector wherein a heterologous nucleotide, e.g. DNA, replaces part or all of the E3 region and/or part or
30 all of the E1 region. The expression system can be used wherein the foreign nucleotide sequences, e.g. DNA, is with or without the control of any other heterologous promoter.

The BAV E1 gene products of the adenovirus
35 of the invention transactivate most of the cellular genes, and therefore, cell lines which constitutively express E1 proteins can express cellular polypeptides at a higher level than normal cell lines. The recombinant mammalian, particularly bovine, cell lines

-30-

of the invention can be used to prepare and isolate polypeptides,, including those such as (a) proteins associated with adenovirus E1A proteins: e.g. p300, retinoblastoma(Rb) protein, cyclins, kinases and the like.; (b) proteins associated with adenovirus E1B protein: e.g. p53 and the like.; (c) growth factors, such as epidermal growth factor (EGF), transforming growth factor (TGF) and the like; (d) receptors such as epidermal growth factor receptor (EGF-R), fibroblast growth factor receptor (FGF-R), tumor necrosis factor receptor (TNF-R), insulin-like growth factor receptor (IFG-R), major histocompatibility complex class I receptor and the like; (e) proteins encoded by proto-oncogenes such as protein kinases (tyrosine-specific protein kinases and protein kinases specific for serine or threonine), p21 proteins (guanine nucleotide-binding proteins with GTPase activity and the like; (f) other cellular proteins such as actins, collagens, fibronectins, integrins, phospholipids, proteoglycans, histones and the like, and (g) proteins involved in regulation of transcription such as TATA-box-binding protein (TBP), TBP-associated factors (TAFs). SP1 binding protein and the like.

25 The invention also includes a method for providing gene therapy to a mammal in need thereof to control a gene deficiency which comprises administering to said mammal a live recombinant bovine adenovirus containing a foreign nucleotide sequence encoding a non-defective form of said gene under conditions wherein the recombinant virus vector genome is incorporated into said mammalian genome or is maintained independently and extrachromosomally to provide expression of the required gene in the target organ or tissue. These kinds of techniques are recently being used by those of skill in the art to replace a defective gene or portion thereof. Examples of foreign genes nucleotide sequences or portions thereof that can be incorporated for use in a

-31-

conventional gene therapy include, cystic fibrosis transmembrane conductance regulator gene, human minidystrophin gene, alpha1-antitrypsin gene and the like.

5

Examples

Described below are examples of the present invention. These examples are provided only for illustrative purposes and are not intended to limit the scope of the present invention in any way. In light of the present disclosure, numerous embodiments within the scope of the claims will be apparent to those of ordinary skill in the art. The contents of the references cited in the specification are incorporated by reference herein.

15

Cells and viruses

Cell culture media and reagents were obtained from GIBCO/BRL Canada (Burlington, Ontario, Canada). Media were supplemented with 25 mM Hepes and 50 µg/ml gentamicin. MDBK cells or MDBK cells transformed with a plasmid containing BAV3 E1 sequences were grown in MEM supplemented with 10% Fetal bovine serum. The wild-type BAV3 ((strain WBR-1) (Darbyshire et al, 1965 J. Comparative Pathology 75:327) was kindly provided by Dr. B. Darbyshire, University of Guelph, Guelph, Canada) and BAV3-luciferase recombinants working stocks and virus titrations were done in MDBK cells.

25
30

Enzymes, bacteria and plasmids

Restriction endonucleases, polymerase chain reaction (PCR) and other enzymes required for DNA manipulations were purchased from Pharmacia LKB Biotechnology (Canada) Ltd. (Dorval, Quebec, Canada), Boehringer-Mannheim, Inc. (Laval or Montreal, Quebec, Canada), New England BioLabs (Beverly, MA), or GIBCO/BRL Canada (Burlington, Ontario, Canada) and used as per manufacturer's instructions. Restriction

35

-32-

enzyme fragments of BAV3 DNA were inserted into pUC18 or pUC19 (Yanich-Penon et al (1985) Gene 33:103-109) following standard procedures (Sambrook et al (1989) Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbour Laboratory, New York). *E. coli* strain DH5 (*supE44 hsdR17 recA1 endA1 gyrA96 thi-1 relA1*) was transformed with recombinant plasmids by electroporation (Dower et al. (1988) Nuc. Acids Res., 16:6127-6145). Plasmid DNA was prepared using the alkaline lysis procedure (Bernboim and Doly (1978) Nuc. Acids Res., 7:1513-1523). The plasmid, pSVOA/L containing the entire cDNA encoding firefly luciferase (de Wet et al (1987) Mol. Cell. Biol. 7:725-737), was a gift from D.R. Helinski, University of California, San Diego, La Jolla, CA.

Construction of recombinant BAV3

MDBK cells transformed with a plasmid containing BAV3 E1 sequences were cotransfected with the wt BAV3 DNA digested with PvuI and the plasmid, pSM51-Luc (Figs. 9 and 10) using the lipofection-mediated cotransfection protocol (GIBCO/BRL, Life Technologies, Inc., Grand Island, NY). The virus plaques produced following cotransfection were isolated, plaque purified and the presence of the luciferase gene in the BAV3 genome was detected by agarose gel electrophoresis of recombinant virus DNA digested with appropriate restriction enzymes.

Southern blot and hybridization

Mock or virus-infected MDBK cells were harvested in lysis buffer (500 µg/ml pronase in 0.01 M Tris, pH 7.4, 0.01 M EDTA, 0.5% SDS) and DNA was extracted (Graham et al (1991) Manipulation of adenovirus vectors In: Methods and Molecular Biology, 7:Gene Transfer and Expression Techniques (Eds. Murray and Walker) Humana Press, Clifton, N.J. pp. 109-128). 100 ng DNA was digested either with BamHI, EcoRI or XbaI and resolved on a 1% agarose gel by

-33-

electrophoresis. DNA bands from the agarose gel were transferred to a *GeneScreenPlus*[™] membrane (Du Pont Canada Inc. (NEN Products), Lachine, Quebec, Canada) by the capillary blot procedure (Southern, E.M. (1975) *J. Mol. Biol.* 98:503-517). Probes were labeled with ³²P using an Oligolabeling Kit (Pharmacia LKB Biotechnology (Canada) Ltd., Dorval, Quebec, Canada) and the unincorporated label was removed by passing the labeled probe through a sephadex G-50 column (Sambrook et al (1989) *supra*). Probes were kept in a boiling water bath for 2 min and used in hybridization experiments following *GeneScreenPlus*[™] hybridization protocol. The DNA bands which hybridized with the probe were visualized by autoradiography.

15

Luciferase assays

The protocol was essentially the same as described (Mittal et al (1993) *Virus Res.* 28:67-90). Briefly, MDBK cell monolayers in 25 mm multi-well dishes (Corning Glass Works, Corning, NY) were infected in duplicate either with BAV3-Luc (3.1) or BAV3-Luc (3.2) at a m.o.i. of 50 p.f.u. per cell. At indicated time points post-infection, recombinant virus-infected cell monolayers were washed once with PBS (0.137 M NaCl, 2.7 mM KCl, 8 mM Na₂HPO₄, 1.5 mM KH₂PO₄) and harvested in 1 ml luciferase extraction buffer (100 mM potassium phosphate, pH 7.8, 1 mM dithiothreitol). The cell pellets were resuspended in 200 µl of luciferase extraction buffer and lysed by three cycles of freezing and thawing. The supernatants were assayed for luciferase activity. For the luciferase assay, 20 µl of undiluted or serially diluted cell extract was mixed with 350 µl of luciferase assay buffer (25 mM glycylglycine, pH 7.8, 15 mM MgCl₂, 5 mM ATP) in a 3.5 ml tube (Sarstedt Inc., St-Laurent, Quebec, Canada). Up to 48 tubes can be kept in the luminometer rack and the equipment was programed to inject 100 µl of luciferin solution (1 mM luciferin in 100 mM potassium phosphate buffer, pH

7.8) in the tube present in the luminometer chamber to start the enzyme reaction. The Luminometer (Packard Picolite Luminometer, Packard Instrument Canada, Ltd., Mississauga, Ontario, Canada) used in the present study produced 300 to 450 light units of background count in a 10 sec reaction time. Known amounts of the purified firefly luciferase were used in luciferase assays to calculate the amount of active luciferase present in each sample.

10

Western blotting

Mock or virus-infected MDBK cells were lysed in 1:2 diluted 2X loading buffer (80 mM Tris-HCl, pH 6.8, 0.67 M urea, 25% glycerol, 2.5% SDS, 1 M mercaptoethanol, 0.001% bromophenol blue), boiled for 3 min and then centrifuged to pellet cell debris. Proteins were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) on 0.1% SDS-10% polyacrylamide gels (Laemmli, et al (1970) Nature 227:680-685). After the end of the run, polypeptide bands in the gel were electrophoretically transferred to a nitrocellulose membrane (Bio-Rad Laboratories, Richmond, CA). The membrane was incubated at room temperature for 2 h with 1:4000 diluted rabbit anti-luciferase antibody (Mittal et al (1993) supra). The binding of anti-luciferase antibody to the specific protein band/s on the membrane was detected with 1:5000 diluted horseradish peroxidase conjugated-goat anti-rabbit IgG (Bio-Rad Laboratories, Richmond, CA) and with an ECL Western blotting detection system (Amersham Canada Ltd., Oakville, Ontario).

20
25
30

Example 1 Cloning of BAV3 E1 Region DNA for sequencing

To complement the restriction site (Kurokawa et al, 1978 J. Virol., 28:212-218; Hu et al, 1984 J. Virol. 49:604-608) other restriction enzyme sites in the BAV3 genome were defined. The 8.4 kilobase pair (kb) SalI B fragment which extends from the left end of the genome to approximately 24% was cloned into the

35

-35-

SmaI-SalI sites of pUC18 essentially as described previously (Graham et al, 1989 EMBO Journal 8:2077-2085). Beginning at the left end of the BAV3 genome, the relevant restriction sites used for subsequent subcloning and their approximate positions are: SacI (2%), EcoRI (3.5%), HindIII (5%), SacI (5.5%), SmaI (5.6%) and HindIII (11%). Through the use of appropriate restriction enzymes, the original plasmid was collapsed to contain smaller inserts which could be sequenced using the pUC universal primers. Some fragments were also subcloned in both pUC18 and pUC19 to allow confirmational sequencing in both directions. These procedures, together with the use of twelve different oligonucleotide primers hybridizing with BAV3 sequences, allowed to sequence the BAV3 genome from its left end to the HindIII site at 11%.

To ensure that some features of the sequence obtained were not unique to the initial clone selected for sequencing, two more pUC19 clones were prepared containing the SalI fragment from a completely independent DNA preparation. These clones were used to confirm the original sequence for the region from approximately 3% to 5.5% of the BAV3 genome.

DNA sequencing reactions were based on the chain-termination method (Sanger et al. 1977 PNAS, USA 74:5463-5467) and manual sequencing followed the DNA sequencing protocol described in the Sequenase™ kit produced by US Biochemical. [α -³⁵S]dATPs was obtained from Amersham Canada Ltd. All oligonucleotides used as primers were synthesized by the Central Facility of the Molecular Biology and Biotechnology Institute (MOBIX) at McMaster University, Hamilton, Ontario. The entire region (0 to 11%) of the BAV3 genome was sequenced by at least two independent determinations for each position by automated sequencing on a 373A DNA Sequencer (Applied Biosystems) using Taq-Dye terminators. Over half of the region was further sequenced by manual procedures to confirm overlaps and other regions of interest.

-36-

DNA sequence analysis and protein comparisons were carried out on a MICROGENIE program.

Example 2 Coding Sequences of the BAV3 E1 Region

5 BAV3 genomic DNA, from the left end of the genome to the HindIII site at approximately 11%, was cloned into plasmids and sequenced by a combination of manual and automated sequencing. An examination of the resultant BAV3 E1 genomic sequence (Fig 1)

10 revealed a number of interesting features relevant both to transactivation and to other functions associated with adenovirus E1 proteins. On the basis of open reading frames (ORFs) it was possible to assign potential coding regions analogous to those

15 defined in human Ad5 (HAd5). As shown in Fig 1, ORFs corresponding roughly to the first exon and unique region of HAd5 E1A as well as ORFs corresponding to the 19k and 58k proteins of E1B and the ORF corresponding to protein IX were all defined in this

20 sequence. The open reading frame defining the probable E1A coding region begins at the ATG at nt 606 and continues to a probable splice donor site at position 1215. The first consensus splice acceptor site after this is located after nt 1322 and defines

25 an intron of 107 base pairs with an internal consensus splice branching site at position 1292. The putative BAV3 E1A polypeptide encoded by a message corresponding to these splice sites would have 211 amino acids and a unmodified molecular weight of

30 23,323. The major homology of the protein encoded by this ORF and HAd5 E1A is in the residues corresponding to CR3 (shown in Fig 2). The homology of amino acid sequences on both sides of the putative intron strengthens the assignment of probable splice donor and acceptor sites. The CR3 has been shown to be of

35 prime importance in the transactivation activity of HAd5 E1A gene products. As seen in Fig. 2A the homology of this sequence in the BAV3 protein to the corresponding region of the 289R E1A protein of HAd5

-37-

includes complete conservation of the
CysX₂CysX₁₃CysX₂Cys sequence motif which defines the
metal binding site of this protein (Berg, 1986 Science
232:485-487) as well as conservation of a number of
5 amino acids within this region and within the promoter
binding region as defined by Lillie and Green 1989
Nature 338:39-44).

The only other region of significant
homology between the BAV3 E1A protein and that of HAd5
10 was a stretch of amino acids known to be important in
binding of the cellular Rb protein to the HAd5 E1A
protein (Dyson et al, 1990 J. Virol. 64:1353-1356).
As shown in Fig 2B, this sequence, which is located
between amino acids 120 and 132 in the CR2 region of
15 HAd5 E1A, is found near the amino (N-) terminus of the
BAV3 protein between amino acids 26 and 37.

An open reading frame from the ATG at nt
1476 to the termination signal at 1947 defines a
protein of 157 amino acids with two regions of major
20 homology to the HAd5 E1B 19k protein. As shown in Fig
3 both the BAV3 and the HAd5 proteins have a centrally
located hydrophobic amino acid sequence. The sequence
in BAV3, with substitutions of valine for alanine and
leucine for valine, should result in a somewhat more
25 hydrophobic pocket than the corresponding HAd5 region.
The other portion of HAd5 19k that may be conserved in
the BAV3 protein is the serine rich sequence found
near the N-terminus (residues 20 to 26) in HAd5 19k
and near the C-terminus (residues 136 to 142) in the
30 BAV3 protein (also shown in Fig 3).

On ORF beginning at the ATG at nt 1850 and
terminating at nt 3110 overlaps the preceding BAV3
protein reading frame and thus has the same
relationship to it as does the HAd5 E1B 56k protein to
35 E1B 19k protein. As shown in Fig 4 this BAV3 protein
of 420R and the corresponding HAd5 E1B 56k protein of
496R show considerable sequence homology over their C-
terminal 346 residues. The N-terminal regions of

-38-

these proteins (not depicted in the figure) show no significant homology and differ in overall length.

Following the E1B ORFs, the open reading frame beginning at nt 3200 and ending at the translation terminator TAA at nt 3575 defines a protein of 125R with an unmodified molecular weight of 13,706. As seen in Fig 5 this protein shares some homology with the structural protein IX of HAd5 particularly in N-terminal sequences.

10

Possible Transcription Control Regions in BAV3 E1

The inverted terminal repeats (ITR) at the ends of the BAV3 genome have been shown to extend to 195 nt (Shinagawa et al, 1987 Gene 55:85-93). The GC-rich 3' portion of the ITR contains a number of consensus binding sites for the transcription stimulating protein SP1 (Dyran and Tijan (1983) Cell 35:79-87) and possible consensus sites for the adenovirus transcription factor (ATF) (Lee et al. (1987) Nature 325:368-372) occur at nts 60 and 220. While there are no exact consensus sites for the factors EF-1A (Bruder and Healing (1989) Mol. Cell Biol. 9:5143-5153) or E2F (Kovesdi et al, 1987 PNAS, USA 84:2180-2184) upstream of the ATG at nt 606, there are numerous degenerate sequences which may define the enhancer region comparable to that seen in HAd5 (Hearing and Shenk, 1986 Cell 45:229-236).

The proposed BAV3 E1A coding sequence terminates at a TGA residue at nt 1346 which is located within a 35 base pair sequence which is immediately directly repeated (see Fig 1). Two repeats of this sequence were detected in three independently derived clones for a plaque purified stock of BAV3. The number of direct repeats can vary in any BAV3 population though plaque purification allows for isolation of a relatively homogeneous population of viruses. That direct repeats in the sequences can function as promoter or enhancer elements for E1B transcription is being tested. There

are no strong polyA addition consensus sites between the E1A and the E1B coding sequences and in fact no AATAA sequence is found until after the protein IX coding sequences following E1B. The TATAAA sequence
5 beginning at nt 1453 could function as the proximal promoter for E1B but it is located closer to the ATG at 1476 than is considered usual (McKnight et al, 1982 Science 217:316-322). The TATA sequence located
10 further upstream immediately before the proposed E1A intron sequence also seems inappropriately positioned to serve as a transcription box for the E1B proteins. There are clearly some unique features in this region of the BAV3 genome.

The transcriptional control elements for the
15 protein IX transcription unit are conventional and well defined. Almost immediately following the open reading frame for the larger E1B protein there is, at nt 3117, a SP1 binding sequence. This is followed at 3135 by a TATAAAT sequence which could promote a
20 transcript for the protein IX open reading frame beginning at the ATG at 3200 and ending with the TAA at 3575. One polyA addition sequence begins within the translation termination codon and four other AATAA sequences are located at nts 3612, 3664, 3796 and
25 3932.

In keeping with the general organization of the E1A region of other adenoviruses, the BAV3 E1A region contains an intron sequence with translation termination codons in all three reading frames and
30 which is therefore probably deleted by splicing from all E1A mRNA transcripts. The largest possible protein produced from the BAV3 E1A region will have 211 amino acid residues and is the equivalent of the 289 amino acid protein translated from the 13s mRNA of
35 HAd5. Two striking features in a comparison of these proteins are the high degree of homology in a region corresponding to CR3 and the absence in BAV3 of most of amino acids corresponding to the second exon of HAd5. In fact the only amino acids encoded in the

-40-

second exon of BAV3 are, those which are considered to constitute part of CR3. A great deal of work carried out with HAd5 has identified the importance of the CR3 sequences in transactivation of other HAd5 genes.

5 While a detailed analysis of the corresponding BAV3 region and its possible role in transactivation of BAV3 genes needs to be carried out, it is none-the-less interesting to note a couple of possibly pertinent features. The HAd5 CR3 region has been
10 operationally subdivided into three regions (Lillie et al, 1989 Nature 338:39-44; see Fig 8); an N-terminal region from 139 to 153 which has four acidic residues and is thought to be important in transcription activation, a central, metal-binding, region defined
15 by the Cys-X₂-Cys-X₁₃-CysX₂-Cys sequence which is essential for both promoter binding and activation, and a C-terminal region (residues 175-189) which is essential for promoter binding. Since, in most instances, E1A protein is thought not to interact
20 directly with DNA (Ferguson et al 1985), the promoter binding regions may be involved in forming associations with proteins which then allow association with DNA. In Fig 2a the BAV3 E1A protein contains the central, metal binding domain and has
25 considerable homology in the carboxy portion of this region. The BAV3 E1A protein also shows identity of sequence with HAd5 in the carboxy 6 amino acids of the promoter binding domain. These features may allow the BAV3 E1A protein to interact with the same
30 transcription activating factors required for HAd5 E1A function. In contrast, except for a Glu-Glu pair there is little homology between the bovine and human viruses in the activation domain. The fact that this domain can be functionally substituted by a
35 heterologous acidic activation sequence (Lillie et al, 1989 supra) suggests that protein specificity is not required in this region and this may allow the BAV3 E1A protein to function in the activation of BAV3 genes. The BAV3 E1A activation region contains six

-41-

acidic residues in the 18 residues amino to the metal binding domain.

The other interesting feature of BAV3 E1A, which is undoubtedly relevant to the oncogenic potential of this virus, is the presence of the sequence Asp27-Leu-Glu-Cys-His-Glu which conforms to, a core sequence known to be important in the binding of cellular Rb and related proteins by the transforming proteins of a number of DNA tumour viruses (Dyson et al, 1990 supra). From deletion mutant analysis there is a clear association between the potential of HAd5 E1A proteins to bind Rb and the ability of the protein to induce morphological transformation in appropriate cells (see references in Dyson et al, 1990 supra). The BAV3 E1A protein is distinct from its HAd5 counterpart in the relative position of this Rb binding sequence which is in the CR2 of HAd5 E1A and near the N-terminus of the BAV3 E1A protein.

Through the use of alternative splice sites HAd5 E1A transcripts can give rise to at least 5 distinct mRNA species (Berk et al, 1978 Cell 14:695-711; Stephens et al, 1987 EMBO Journal 6:2027-2035). Whether BAV3, like HAd5, can generate a number of different mRNA species through the use of alternative splice sites in the E1A transcripts remains to be determined. For example a potential splice donor site which could delete the sequence equivalent to the unique sequence of HAd5 is present immediately after nt 1080 but it is not known if this site is actually used.

HAd5 E1B encodes two proteins (19k and 56k) either of which can cooperate with E1A, by pathways which are additive and therefore presumably independent (McLorie et al, 1991 J. Gen. Virol. 72:1467-1471), to produce morphological transformation of cells in culture (see for example: Branton et al, 1985 supra; Graham, 1984 supra). The significance of the conservation of the hydrophobic stretch of amino

-42-

acids in the central portion of the shorter E1B proteins of HAd5 and BAV3 is not clear as yet. A second short region of homology Gln-Ser-Ser-X-Ser-Thr-Ser at residue 136 near the C-terminus of the BAV3 protein is located near the N-terminus at residue 20 in the HAd5 19k protein. The major difference in both length and sequence of the larger (420R) E1B protein of BAV3 from the corresponding HAd5 protein (496R) is confined to the N-terminus of these proteins. The two proteins show considerable evolutionary homology in the 345 amino acids that extend to their C-termini. A similar degree of homology extends into the N-terminal halves of protein IX of BAV3 and HAd5. Taken together these analyses suggest that while BAV3 and the human adenoviruses have diverged by simple point mutational events in some regions, more dramatic genetic events such as deletion and recombination may have been operating in other regions particularly those defining the junction between E1A and E1B.

20

Example 3 Cloning and sequencing of the BAV3 E3 and fibre genes

The general organization of adenovirus genomes seems to be relatively well conserved so it was possible to predict, from the locations of a number of HAd E3 regions, that BAV E3 should lie between map units (m.u.) 77 to 86. To prepare DNA for cloning and sequencing, BAV3 (strain WBR-1) was grown in Madin-Darby bovine kidney (MDBK) cells, virions were purified and DNA was extracted (Graham, F.L. & Prevec, L. (1991) Methods in Molecular Biology, vol. 7, Gene Transfer and Expression Protocols, pp. 109-146. Edited by E.J. Murray, Clifton, New Jersey; Humana Press.). Previously published restriction maps for *EcoRI* and *BamHI* (Kurokawa et al., 1978) were confirmed (Fig. 6). The *BamHI* D and *EcoRI* F fragments of BAV3 DNA were isolated and inserted into pUC18 and pUC19 vectors, and nested sets of deletions were made using exonuclease III and S1 nuclease (Henikoff, S.

-43-

(1984) Gene, 28:351-359). The resulting clones were sequenced by the dideoxynucleotide chain termination technique (Sanger, F., Nicklen, S. & Coulson, A.R. (1977) Proceedings of the National Academy of Sciences, U.S.A., 74:5463-5467). The nucleotide sequence from positions 1 to 287 was obtained from the right end of the *Bam*HI B fragment (Fig. 6). The sequence of the regions spanning (i) the *Bam*HI site at nucleotide 3306 and the *Eco*RI site at nucleotide 3406, and (ii) the *Eco*RI site at nucleotide 4801 and the nucleotide 5100 was obtained from a plasmid containing the *Xba*I C fragment (m.u. 83 to 100; not shown) using primers hybridizing to BAV3 sequences. Analysis of the sequence was performed with the aid of the PC/GENE sequence analysis package developed by Amos Bairoch, Department of Medical Biochemistry, University of Geneva, Switzerland.

The 5100 nucleotide sequence which extends between 77 and 92 m.u. of the BAV3 genome is shown in Fig. 7. The upper strand contains 14 open reading frames (ORFs) which could encode polypeptides of 60 amino acid residues or more (Fig. 6 and 7). The lower strand contains no ORF encoding a protein of longer than 50 amino acids after an initiation codon. The predicted amino acid sequence for each ORF on the upper strand was analyzed for homology with predicted amino acid sequences from several sequenced Ads: HAd2 (Hérissé, J., Courtois, G. & Galibert, F. (1980) Nucleic Acids Research, 8:2173-2192; Hérissé, J., Courtois, G. & Galibert, F. (1981) Nucleic Acids Research, 9:1229-1249), -3 (Signas, C., Akusjarvi, G. & Pettersson, U. (1985) Journal of Virology, 53:672-678.), -5 (Cladaras, C. & Wold, W.S.M. (1985) Virology, 140:28-43), -7 (Hong, J.S., Mullis, K.G. & Engler, J.A. (1988) Virology, 167:545-553) and -35 (Flomenberg, P.R., Chen, M. & Horwitz, M.S. (1988) Journal of Virology, 62:4431-4437), and murine Ad1 (MAd1) (Raviprakash, K.S., Grunhaus, A., El Kholy, M.A. & Horwitz, M.S. (1989) Journal of Virology, 63:5455-

-44-

5458) and canine Ad1 (CAD1) (Dragulev, B.P., Sira, S., Abouhaidar, M.G. & Campbell, J.B. (1991) Virology, 183:298-305). Three of the BAV3 ORFs exhibited homology with characterized HAd proteins pVIII, fibre and the 14.7K E3 protein. The amino acid sequence predicted from BAV3 ORF 1 shows overall identity of approximately 55% when compared to the C-terminal 75% of HAd2 pVIII (Cladaras & Wold, 1985, supra) (Fig. 8a), indicating that ORF 1 encodes the right end of BAD3 pVIII. Near the C-terminal end of BAD3 pVIII there is a 67 amino acid stretch (residues 59 to 125; Fig. 8a) which has 75% identity with HAd2 pVIII. This region has previously been shown to be highly conserved among different Ads (Cladaras & Wold, 1985, supra; Signas, C., Akusjarvi, G. & Pettersson, U. (1986) Gene, 50:173-184,; Raviprakash et al., 1989, supra; Dragulev et al., 1991, supra).

The fibre protein is present on the surface of the virion as long projections from each vertex of the icosahedral capsid and is involved in a number of Ad functions including attachment of the virus to the cell surface during infection, assembly of virions and antigenicity (Philipson, L. (1983) Current Topics in Microbiology and Immunology, 109:1-52). On the basis of the primary structure of HAd2 fibre protein, it has been proposed that the shaft region (between amino acid residues 40 and 400) is composed of a number of repeating structural motifs containing about 15 hydrophobic residues organized in two short β -sheets and two β -bends (Green, N.M., Wrigley, N.G., Russell, W.C., Martin, S.R. & McLachlan, A.D. (1983) EMBO Journal, 2:1357-1365). The amino acid sequences at the N terminus of the BAV3 ORF 6-encoded protein share about 60% identity with the HAd2 fibre protein tail, but there is little or no similarity in the knob region, and about 45% identity overall (Fig. 8c). The BAD3 fibre gene would encode a protein of 976 residues if no splicing occurs, i.e. 394 amino acid residues longer than the HAd2 fibre protein. The number of

repeating motifs in the shaft region of the fibre protein from different Ads varies between 28 and 23 (Signas et al., 1985, supra; Chroboczek, J. & Jacrot, B. (1987) *Virology*, 161:549-554; Hong et al., 1988, supra; Raviprakash et al., 1989, supra; Dragulev et al., 1991, supra). The BAV3 fibre protein can be organized into 52 such repeats in this region (not shown), which would account for most of the difference in size compared to those of HAd2, HAd3, HAd5, HAd7, CAd1 and MAd1 (Signas et al., 1985, supra; Hérissé et al., 1980, supra; Hérissé & Galibert, 1981, supra; Hong et al., 1988, supra; Raviprakash et al., 1989, supra; Dragulev et al., 1991, supra).

HAd2 and HAd5 E3 lies between the pVIII and the fibre genes and encodes at least 10 polypeptides (Cladaras & Wold, 1985, supra). The promoter for E3 of these two serotypes lies within the sequences encoding pVIII, about 320 bp 5' of the termination codon. No consensus TATA box is found in the corresponding region of the BAV3 sequences. A non-canonical polyadenylation signal (ATAAA) for E3 transcripts is located at position 1723, between the end of the putative E3 region and the beginning of ORF 6, encoding the fibre protein, and two consensus signals are located within ORF 6 at positions 2575 and 3565. The polyadenylation signal for the fibre protein is located at nucleotide 4877. Six ORFs were identified in the BAV3 genome between the pVIII and the fibre genes, but only four (ORFs 2, 3, 4 and 5) have the potential to encode polypeptides of at least 50 amino acids after an initiation codon (Fig. 7). The amino acid sequence predicted to be encoded by ORF 2 is 307 residues long and contains eight potential N-glycosylation sites (Fig. 7) as well as a hydrophobic sequence which may be a potential transmembrane domain (PLLFAFVLCTGCAVLLTAFGPSILSGT) between residues 262 and 289. This domain may be a part of the protein homologous to the HAd2 and HAd5 19K E3 glycoprotein (Cladaras & Wold, 1985, supra), and the proposed CAd1

-46-

22.2K protein (Dragulev et al., 1991, supra), but ORF 2 does not show appreciable homology with these proteins. The ORF 4 shows approximately 44% identity with the 14.7K E3 protein of HAd5 (Fig. 6 and 8b), which has been shown to prevent lysis of virus-infected mouse cells by tumour necrosis factor (Gooding, L.R., Elmore, L.W., Tollefson, A.E., Brody, H.A. & Wold, W.S.M. (1988) Cell, 53:341-346; Wold, W.S.M. & Gooding, L.R. (1989) Molecular Biology and Medicine, 6:433-452). Analysis of the 14.7K protein sequence from HAd2, -3, -5 and -7 has revealed a highly conserved domain, which in HAd5 lies between amino acid residues 41 and 56 (Horton, T.M., Tollefson, A.E., Wold, W.S.M. & Gooding, L.R. (1990) Journal of Virology, 64:1250-1255). The corresponding region in the BAV3 ORF 4-encoded protein, between amino acids 70 and 85, contains 11 amino acids identical to those of the HAd5 14.7K protein conserved domain (Fig. 8b).

The BAV3 E3 region appears to be approximately 1.5kbp long, about half the size of those of HAd2 and -5 (Cladaras & Wold, 1985, supra), and novel splicing events in BAV3 E3 would be required to generate more homologues to the HAd3 E3 proteins. A similarly short E3 region has been reported for MAd1 (RAviprakash et al., 1989, supra) and CAd1 (Dragulev et al., 1991, supra).

Example 4 Construction of BAV3-luciferase recombinants

Adenovirus-based mammalian cell expression vectors have gained tremendous importance in the last few years as a vehicle for recombinant vaccine delivery, and also in gene therapy. BAV3-based expression vectors have a greater potential for developing novel recombinant vaccines for veterinary use. To show that BAV3 E3 gene products are not essential for virus growth in cultured cells and this locus could be used to insert foreign DNA sequences, a

-47-

1.7 kb fragment containing the firefly luciferase gene was introduced in the 696 bp deletion of the E3 region of the BAV3 genome in the E3 parallel orientation to generate a BAV3 recombinant.

5 The rationale of using the luciferase gene is that it acted as a highly sensitive reporter gene when introduced in the E3 region of the HAd5 genome to generate HAd5-Luc recombinants (Mittal et al (1993) Virus Res. 28:67-90).

10 To facilitate the insertion of the firefly luciferase gene into the E3 region of the BAV3 genome, a BAV3 E3 transfer vector containing the luciferase gene was constructed (Fig. 9). The BAV3 E3 region falls approximately between m.u. 77 and 82. In
15 our first series of vectors we replaced a 696 bp XhoI-NcoI E3 deletion (between m.u. 78.8 and 80.8) with a NruI-SalI cloning sites for insertion of foreign genes to obtain pSM14del2. A 1716 bp BsmI-SspI fragment
20 containing the luciferase gene was isolated and first inserted into an intermediate plasmid, pSM41, in the E3 locus at the SalI site by blunt end ligation to generate pSM41-Luc. The luciferase gene without any
25 exogenous regulatory sequences, was inserted into the E3 locus in the same orientation as the E3 transcription unit. The kan^r gene was inserted into pSM41-Luc at the XbaI site present within the
luciferase gene to generate an amp^r/kan^r plasmid, pSM41-Luc-Kan. A 7.7 kb fragment containing the BAV3
30 sequences along with the luciferase gene and the kan^r gene was obtained from pSM41-Luc-Kan by digestion with BamHI and inserted into an amp^r plasmid, pSM51
partially digested with BamHI to replace a 3.0 kb BamHI fragment (lies between m.u. 77.8 and 86.4) to
35 generate a doubly resistant (kan^r & amp^r) plasmid, pSM51-Luc-Kan. The kan^r gene was deleted from pSM51-Luc-Kan by partial cleavage with XbaI to generate
pSM51-Luc containing the luciferase gene in the E3-parallel orientation.

-48-

MDBK cells transformed with a plasmid containing the BAV3 E1 sequences was cotransfected with the wt BAV3 DNA digested with PvuI, which make two cuts within the BAV3 genome at m.u 65.7 and 71.1, and the plasmid, pSM51-Luc to rescue the luciferase gene in E3 of the BAV3 genome by *in vivo* recombination (Fig. 10). The digestion of the wt BAV3 DNA with PvuI was helpful in minimizing the generation of the wt virus plaques following cotransfection. The left end of the wt BAV3 genome represented by PvuI 'A' fragment falls between m.u. 0 and 65.7, and pSM51-Luc which extends between m.u. 31.5 and 100 (except for E3 deletion replaced with the luciferase gene) have sufficient overlapping BAV3 DNA sequences to generate recombinant viruses.

Two virus plaques were obtained in two independent cotransfection experiments which were grown in MDBK cells. The viral DNA from both plaques was extracted and analyzed by agarose gel electrophoresis after digesting either with BamHI, EcoRI or XbaI to identify the presence and orientation of the luciferase gene in the viral genome (data not shown). In the genomes of both recombinants, the luciferase gene was present in the E3 region in the E3 parallel orientation. The BAV3-luciferase recombinants were plaque purified and named BAV3-Luc (3.1) and BAV3-Luc (3.2) to represent plaques obtained from two independent experiments. Since both recombinant virus isolates were identical they will be referred to as BAV3-Luc. The presence of the luciferase gene in BAV3-Luc isolates are further confirmed by Southern blot analyses and luciferase assays using extracts from recombinant virus-infected cells.

Characterization of BAV3-recombinants

Southern blot analyses of the wt BAV3 and recombinants genomic DNA digested either with BamHI, EcoRI or XbaI, were carried out to confirm the

-49-

presence and orientation of the luciferase gene in the E3 locus and the deletion of the 696 bp XhoI-NcoI fragment from E3 of the BAV3-Luc genome (Fig. 11). When the blot was probed with a 696 XhoI-NcoI fragment of E3 of the BAV3 genome (panel A, lanes 4 to 9) no hybridization signal was detected with the DNA fragments from the recombinant viruses, however, the expected bands (3.0 kb BamHI, 8.1 kb EcoRI, and 18.5 kb XbaI) of the wt BAV3 DNA fragments (panel A, lanes 10 to 12) showed hybridization, confirming that the 696 bp XhoI-NcoI fragment of the E3 region was indeed deleted in the BAV3-Luc genomic DNA. In panel B, when an identical blot was probed with the luciferase gene, there were strong hybridization signals with the DNA fragments from the recombinant viruses (4.0 kb BamHI (lane 4 & 7), 6.0 kb & 3.2 kb EcoRI (lanes 5 & 8), 16.7 kb & 2.9 kb XbaI (lanes 6 & 9)). These results confirmed that the BAV3-Luc contains the luciferase gene in the E3 parallel orientation with a 696 bp XhoI-NcoI E3 deletion.

The growth characteristics of the recombinant viruses was compared with the wt BAV3 in a single step growth curve (Fig. 12). Virus titers in MDBK cells-infected with the wt BAV3 started increasing at 12 h post-infection reaching a maximum at 36-48 h post-infection and then declined thereafter. Virus titers of the recombinant viruses also started increasing at 12 h postinfection reaching a maximum at 48 h post-infection and then declined, however, the titers of recombinant viruses remained approximately one log lower than the wt virus. The plaque size of the recombinant viruses were also comparatively smaller than the wt virus (data not shown).

Kinetics of luciferase expression by BAV3-Luc

Luciferase activity in BAV3-Luc-infected MDBK cells was monitored at different times post-infection by luciferase assays (Fig. 13). A low level

of luciferase activity was first observed at 12 h post-infection reaching a peak at 30 h post-infection and then dropped subsequently. At 30 h post-infection, approximately 425 pg luciferase was
5 detected in 4×10^5 BAV3-Luc (3.1)-infected MDBK cells. In MDBK cells-infected with the wt BAV3, luciferase expression was not detected (data not shown). The kinetics of luciferase expression by BAV3-Luc (3.1) and BAV3-Luc (3.2) appears very much similar. The
10 kinetics of luciferase expression also showed that the majority of enzyme expression in virus-infected cells seemed to occur late in infection. To determine luciferase expression in the absence of viral DNA replication, BAV3-Luc-infected MDBK cells were
15 incubated in the presence of an inhibitor of DNA synthesis, 1- β -D-arabinofuranosyl cytosine (AraC) and luciferase activity was measured in virus-infected cell extracts at various times post-infection and compared to luciferase expression obtained in the
20 absence of AraC (Fig. 14). When the recombinant virus-infected cells were incubated in the presence of AraC, luciferase expression at 18, 24 and 30 h post-infection was approximately 20-30% of the value obtained in the absence of AraC. These results
25 indicated that the majority of luciferase expression in MDBK cells infected with BAV3-Luc took place after the onset of viral DNA synthesis. To confirm this MDBK cells-infected with the BAV3-Luc were grown in the absence or presence of AraC, harvested at 18 h, 24
30 h, and 30 h post-infection, viral DNA extracted and analyzed by dot blot analysis using pSM51-Luc (see Fig. 9) as a probe (data not shown). In the presence of AraC, viral DNA synthesis was severely reduced compared to viral DNA synthesis in the absence of
35 AraC.

Western blot analysis of BAV3-Luc-infected cells

Luciferase was expressed as an active enzyme as determined by luciferase assays using extracts from

-51-

MDBK cells-infected with BAV3-Luc (see Fig. 13). The luciferase gene without any exogenous regulatory sequences was inserted into E3 of the BAV3 genome, therefore, there was a possibility of luciferase expression as a fusion protein with part of an E3 protein if the luciferase gene was in the same frame, Such as, F1 and F3 which represent open reading frames (ORFs) for E3 proteins (Fig. 15) or the fusion protein may arise due to recognition of an upstream initiation codon in the luciferase ORF. To explore this possibility we sequenced the DNA at the junction of the luciferase gene and the BAV3 sequences with the help of a plasmid, pSM51-Luc and a synthetic primer design to bind luciferase coding sequences near the initiation codon (data not shown). The luciferase coding region fell in frame F2. The luciferase initiation codon was the first start codon in this frame, however, the ORF started at 84 nucleotides upstream of the luciferase start codon. To further confirm that luciferase protein is of the same molecular weight as purified firefly luciferase, unlabeled mock-infected, wt BAV3-infected or BAV3-Luc-infected MDBK cell extracts were reacted with an anti-luciferase antibody in a Western blot (Fig. 16). A 62 kDa polypeptide band was visible in the BAV3-Luc (lane 3 and 4)-infected cell extracts which were of the same molecular weight as pure firefly luciferase (lane 5). We are not sure whether a band of approximately 30 kDa which also reacted with the anti-luciferase antibody in lanes 3 and 4 represented a degraded luciferase protein.

The majority of luciferase expression is probably driven from the major late promoter (MLP) to provide expression paralleling viral late gene expression, moreover, the enzyme expression seen in the presence of AraC may be taking place from the E3 promoter. In HAd5 vectors, foreign genes without any exogenous regulatory sequences when inserted in E3 also displayed late kinetics and were inhibited by

-52-

AraC. The BAV3 recombinant virus replicated relatively well in cultured cells but not as good as the wt BAV3. This is not surprising as infectious virus titers of a number of HAd5 recombinants were slightly lower than the wt HAd5 (Bett et al (1993) J. Virol. 67:5911-5921). This may be because of reduced expression of fiber protein in recombinant adenoviruses having inserts in the E3 region compared to the wt virus (Bett et al, supra and Mittal et al (1993) Virus Res. 28:67-90).

The E3 of BAV3 is approximately half the size of the E3 region of HAd2 or HAd5 and thus has the coding potential for only half the number of proteins compared to E3 of HAd2 or HAd5 (Cladaras et al (1985) Virology 140:28-43; Herisse et al (1980) Nuc. Acids Res. 8:2173-2192; Herisse et al (1981) Nuc. Acids Res. 9:1229-1249 and Mittal et al (1993) J. Gen. Virol. 73:3295-3000). BAV3 E3 gene products have been shown to be not required for virus growth in tissue culture. However, presently it is known that BAV3 E3 gene products also evade immune surveillance in vivo like HAd5 E3 proteins. One of the BAV3 E3 open reading frames (ORFs) has been shown to have amino acid homology with the 14.7 kDa E3 protein of HAd5 (Mittal et al (1993) supra). The 14.7 kDa E3 protein of HAd5 prevents lysis of virus-infected mouse cells by tumour necrosis factor (Gooding et al (1988) Cell 53:341-346 and Horton et al (1990) J. Virol. 64:1250-1255). The study of pathogenesis and immune responses of a series of BAV3 E3 deletion mutants in cattle provides very useful information regarding the role of E3 gene products in modulating immune responses in their natural host.

The BAV3-based vector has a 0.7 kb E3 deletion which can hold an insert up to 2.5 kb in size. The BAV3 E3 deletion can extend probably up to 1.4 kb which in turn would also increase the insertion capacity of this system. The role of the MLP and the E3 promoter is examined to determine their ability to

drive expression of a foreign gene inserted into E3 when a proper polyadenylation signal is provided. Exogenous promoters, such as, the simian virus 40 (SV40) promoter (Subramant et al (1983) Anal. Biochem. 135:1-15), the human cytomegalovirus immediate early promoter (Boshart et al (1985) Cell 43:215-222), and the human beta-actin promoter (Gunning et al (1987) PNAS, USA 84:4831-4835) are tested to evaluate their ability to facilitate expression of foreign genes when introduced into E3 of the BAV3 genome.

Recently HAd-based expression vectors are under close scrutiny for their potential use in human gene therapy (Ragot et al (1993) Nature 361:647-650; Rosenfeld et al (1991) Science 252:431-434; Rosenfeld et al (1992) Cell 68:141-155 and Stratford-Perricaudet et al (1990) Hum. Gene. Ther. 1:241-256). A preferable adenovirus vector for gene therapy would be one which maintains expression of the required gene for indefinite or for a long period in the target organ or tissue. It may be obtained if the recombinant virus vector genome is incorporate into the host genome or maintained its independent existence extrachromosomally without active virus replication. HAds replicate very well in human, being their natural host. HAds can be made defective in replication by deleting the E1 region, however, how such vectors would maintain the expression of the target gene in a required fashion is not very clear. Moreover, the presence of anti-HAds antibodies in almost every human being may create some problems with the HAd-based delivery system. The adenovirus genomes have a tendency to form circles in non-permissive cells. BAV-based vectors could provide a possible alternative to HAd-based vectors for human gene therapy. As BAV3 does not replicate in human, the recombinant BAV3 genomes may be maintained as independent circles in human cells providing expression of the essential protein for a long period of time.

-54-

The foreign gene insertion in animal adenoviruses is much more difficult than HAdS because it is hard to develop a cell line which is also good for adenovirus DNA-mediated transfection. This may be one of the major reasons that the development of an animal adenovirus-based expression system has not been reported so far. It took us more than a year to isolate a cell line suitable for BAV3 DNA-mediated transfection. However, the rapid implementation of BAV-based expression vectors for the production of live virus recombinant vaccines for farm animals, is very promising. BAVs grow in the respiratory and gastrointestinal tracts of cattle, therefore, recombinant BAV-based vaccines have use to provide a protective mucosal immune response, in addition to humoral and cellular immune responses, against pathogens where mucosal immunity plays a major role in protection.

Example 5 Generation of cell lines transformed with the BAV3 E1 sequences

MDBK cells in monolayer cultures were transfected with pSM71-neo, pSM61-kan1 or pSM61-kan2 by a lipofection-mediated transfection technique (GIBCO/BRL, Life Technologies, Inc., Grand Island, NY). At 48 h after transfection, cells were maintained in the MEM supplemented with 5% fetal bovine serum and 700 µg/ml G418. The medium was changed every 3rd day. In the presence of G418, only those cells would grow which have stably incorporated the plasmid DNA used in transfection experiments into their genomes and are expressing the neo^r gene. The cells which have incorporated the neo^r gene might also have taken up the BAV3 E1 sequences and thus expressing BAV3 E1 protein/s. A number of neo^r (i.e., G418-resistant) colonies were isolated, expended and tested for the presence of BAV3 E1 message/s by Northern blot analyses using a DNA probe containing only the BAV3 E1 sequences. Expression of BAV3 E1

-55-

protein/s were confirmed by a complimentation assay using a HAd5 deletion mutant defective in E1 function due to an E1 deletion.

5 Fetal bovine kidney cells in monolayers were also transfected with pSM71-neo, pSM61kan-1 or pSM61-kan2 by the lipofection-mediated transfection technique, electroporation (Chu et al (1987) Nucl. Acids Res. 15:1311-1326), or calcium phosphate precipitation technique (Graham et al (1973) Virology 10 52:456-467). Similarly, a number of G418-resistant colonies were isolated, expended and tested for the presence of BAV3 E1 gene products as mentioned above.

15 Example 6 Generation of a BAV3 recombinant containing the beta-galactosidase gene as an E1 insert

As E1 gene products are essential for virus replication, adenovirus recombinants containing E1 inserts will grow only in a cell line which is transformed with the adenovirus E1 sequences and expresses E1. A number of cell line which are transformed with the BAV3 E1 sequences were isolated as described earlier. The technique of foreign gene insertions into the E1 regions is similar to the gene insertion into the E3 region of the BAV3 genome, 25 however, for insertion into E1 there is a need of an E1 transfer plasmid which contains DNA sequences from the left end of the BAV3 genome, an appropriate deletion and a cloning site for the insertion of foreign DNA sequences. G418-resistant MDBK cell 30 monolayers were cotransfected with the wild-type (wt) BAV3 DNA and pSM71-Z following the lipofection-mediated transfection procedure (GIBCO/BRL, Life Technologies, Inc., Grand Island, NY). The monolayers were incubated at 37°C under an agarose overlay. 35 After a week post-incubation an another layer of overlay containing 300 ug/ml Blu-gal™ (GIBCO/BRL Canada, Burlington, Ontario, Canada) was put onto each monolayer. The blue plaques were isolated, plaque purified and the presence of the beta-galactosidase

-56-

gene in the BAV3 genome was identified by agarose gel electrophoresis of recombinant virus DNA digested with suitable restriction enzymes and confirmed by beta-galactosidase assays using extracts from recombinant virus infected cells.

Deposit of Biological Materials

The following materials were deposited and are maintained with the Veterinary Infectious Disease Organization (VIDO), Saskatoon, Saskatchewan, Canada.

The nucleotide sequences of the deposited materials are incorporated by reference herein, as well as the sequences of the polypeptides encoded thereby. In the event of any discrepancy between a sequence expressly disclosed herein and a deposited sequence, the deposited sequence is controlling.

	<u>Material</u>	<u>Internal Accession No.</u>	<u>Deposit</u>
	<u>Date</u>		
	<u>Recombinant plasmids</u>		
20	psM51	psM51	Dec 6, 1993
	psM71	psM71	Dec 6, 1993
	<u>Recombinant cell lines</u>		
	MDBK cells transformed with BAV3 E1 sequences(MDBK-BAVE1)		Dec 6, 1993
25	Fetal bovine kidney cells transformed with BAV3 E1 sequences(FBK-BAV-E1)		Dec 6, 1993

While the present invention has been illustrated above by certain specific embodiments, the specific examples are not intended to limit the scope of the invention as described in the appended claims.

35

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: UNIVERSITY OF SASKATCHEWAN
- (ii) TITLE OF INVENTION: RECOMBINANT PROTEIN PRODUCTION IN BOVINE ADENOVIRUS EXPRESSION VECTOR SYSTEM
- (iii) NUMBER OF SEQUENCES: 34
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: SCOTT & AYLEN
 - (B) STREET: 60 QUEEN STREET
 - (C) CITY: OTTAWA
 - (D) PROVINCE: ONTARIO
 - (E) COUNTRY: CANADA
 - (F) POSTAL CODE: K1P 5Y7
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: JOAN M. VAN ZANT
 - (B) REFERENCE/DOCKET NUMBER: PAT 21976TW-90
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 1-416-368-2400
 - (B) TELEFAX: 1-416-363-7246

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 4060 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: join(606..1215, 1323..1345)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
CATCATCAAT AATCTACAGT AACTGATGG CAGCGGTCCA ACTGCCAATC ATTTTGGCCA    60
CGTCATTAT GACGCAACGA CGGCGAGCGT GCGTGCTGA CGTAACTGTG GGGCGGAGCG    120
CGTCGCGGAG GCGGCGGCGC TGGGCGGGGC TGAGGCGGCG GGGGCGGCG CGCGGGGCGG    180
CGCGCGGGGC GGGGCGAGGG GCGGAGTTCC GCACCCGCTA CGTCATTTTC AGACATTTT    240
TAGCAAATTT GCGCCTTTTG CAAGCATTTT TCTCACATTT CAGGTATTTA GAGGGCGGAT    300
TTTTGGTGTT CGTACTTCCG TGTCACATAG TTTACTGTCA ATCTTCATTA CGGCTTAGAC    360
AAATTTTCGG CGTCTTTTCC GGGTTTATGT CCCC GGTCAC CTTTATGACT GTGTGAAACA    420
```

-58-

	CACCTGCCCA TTGTTTACCC TTGGTCAGTT TTTTCGTCTC CTAGGGTGGG AACATCAAGA	480
	ACAAATTTGC CGAGTAATTG TGCACCTTTT TCCGCGTTAG GACTGCGTTT CACACGTAGA	540
	CAGACTTTTT CTCATTTTCT CACACTCCGT CGTCCGCTTC AGAGCTCTGC GTCTTCGCTG	600
5	CCACC ATG AAG TAC CTG GTC CTC GTT CTC AAC GAC GGC ATG AGT CGA Met Lys Tyr Leu Val Leu Val Leu Asn Asp Gly Met Ser Arg	647
	ATT GAA AAA GCT CTC CTG TGC AGC GAT GGT GAG GTG GAT TTA GAG TGT Ile Glu Lys Ala Leu Leu Cys Ser Asp Gly Glu Val Asp Leu Glu Cys	695
	CAT GAG GTA CTT CCC CCT TCT CCC GCG CCT GTC CCC GCT TCT GTG TCA His Glu Val Leu Pro Pro Ser Pro Ala Pro Val Pro Ala Ser Val Ser	743
10	CCC GTG AGG AGT CCT CCT CCT CTG TCT CCG GTG TTT CCT CCG TCT CCG Pro Val Arg Ser Pro Pro Pro Leu Ser Pro Val Phe Pro Pro Ser Pro	791
	CCA GCC CCG CTT GTG AAT CCA GAG GCG AGT TCG CTG CTG CAG CAG TAT Pro Ala Pro Leu Val Asn Pro Glu Ala Ser Ser Leu Leu Gln Gln Tyr	839
15	CGG AGA GAG CTG TTA GAG AGG AGC CTG CTC CGA ACG GCC GAA GGT CAG Arg Arg Glu Leu Leu Glu Arg Ser Leu Leu Arg Thr Ala Glu Gly Gln	887
	CAG CGT GCA GTG TGT CCA TGT GAG CGG TTG CCC GTG GAA GAG GAT GAG Gln Arg Ala Val Cys Pro Cys Glu Arg Leu Pro Val Glu Glu Asp Glu	935
	TGT CTG AAT GCC GTA AAT TTG CTG TTT CCT GAT CCC TGG CTA AAT GCA Cys Leu Asn Ala Val Asn Leu Leu Phe Pro Asp Pro Trp Leu Asn Ala	983
20	GCT GAA AAT GGG GGT GAT ATT TTT AAG TCT CCG GCT ATG TCT CCA GAA Ala Glu Asn Gly Gly Asp Ile Phe Lys Ser Pro Ala Met Ser Pro Glu	1031
	CCG TGG ATA GAT TTG TCT AGC TAC GAT AGC GAT GTA GAA GAG GTG ACT Pro Trp Ile Asp Leu Ser Ser Tyr Asp Ser Asp Val Glu Glu Val Thr	1079
25	AGT CAC TTT TTT CTG GAT TGC CCT GAA GAC CCC AGT CGG GAG TGT TCA Ser His Phe Phe Leu Asp Cys Pro Glu Asp Pro Ser Arg Glu Cys Ser	1127
	TCT TGT GGG TTT CAT CAG GCT CAA AGC GGA ATT CCA GGC ATT ATG TGC Ser Cys Gly Phe His Gln Ala Gln Ser Gly Ile Pro Gly Ile Met Cys	1175
	AGT TTG TGC TAC ATG CGC CAA ACC TAC CAT TGC ATC TAT A GTAAGTACAT Ser Leu Cys Tyr Met Arg Gln Thr Tyr His Cys Ile Tyr	1225
30	TCTGTAAGG AACATCTTGG TGATTCTAG GTATTGTTA GGGATTAAC TGGTGGAGTG	1285
	ATCTTAATCC GGCATAACCA AATACATGTT TTCACAG GT CCA GTT TCT GAA GAG Ser Pro Val Ser Glu Glu	1339
	GAA ATG TGAGTCATGT TGACTTTGGC GCGCAAGAGG AAATGTGAGT CATGTTGACT Glu Met	1395
35	TTGGCGCGCC CTACGGTGAC TTTAAGCAA TTTGAGGATC ACTTTTTTGT TAGTCGCTAT	1455
	AAAGTAGTCA CGGAGTCTTC ATGGATCACT TAAGCGTTCT TTTGGATTG AAGCTGCTTC	1515
	GCTCTATCGT AGCGGGGGCT TCAAATCGCA CTGGAGTGTG GAAGAGGCGG CTGTGGCTGG	1575
	GACGCCTGAC TCAACTGGTC CATGATACCT GCGTAGAGAA CGAGAGCATA TTTCTCAATT	1635
	CTCTGCCAGG GAATGAAGCT TTTTAAAGGT TGCTTCGGAG CGGCTATTTT GAAGTGTTTG	1695

	ACGTGTTTGT GGTGCTGAG CTGCATCTGG AACTCCGGG TCGAGTGGTC GCCGCTCTTG	1755
	CTCTGCTGGT GTTCATCCTC AACGATTTAG ACGTAATTC TGCTTCTTCA GGCTTTGATT	1815
	CAGGTTTTCT CGTGACCGT CTCTGCGTGC CGCTATGGCT GAAGGCCAGG GCGTTCAAGA	1875
	TCACCCAGAG CTCCAGGAGC ACTTCGCAGC CTTCCTCGTC GCCCGACAAG ACGACCCAGA	1935
5	CTACCAGCCA GTAGACGGGG ACAGCCCACC CCGGGCTAGC CTGGAGGAGG CTGAACAGAG	1995
	CAGCACTCGT TTCGAGCACA TCAGTTACCG AGACGTGGTG GATGACTTCA ATAGATGCCA	2055
	TGATGTTTTT TATGAGAGGT ACAGTTTTGA GGACATAAAG AGCTACGAGG CTTTGCCTGA	2115
	GGACAATTG GAGCAGCTCA TAGCTATGCA TGCTAAAATC AAGCTGCTGC CCGGTCGGGA	2175
	GTATGAGTTG ACTCAACCTT TGAACATAAC ATCTTGCGCC TATGTGCTCG GAAATGGGGC	2235
10	TACTATTAGG GTAACAGGGG AAGCCTCCCC GGCTATTAGA GTGGGGGCCA TGGCCGTGGG	2295
	TCCGTGTGTA ACAGGAATGA CTGGGGTGAC TTTTGTGAAT TGTAGGTTTG AGAGAGAGTC	2355
	AACAATTAGG GGGTCCCTGA TACGAGCTTC AACTCACGTG CTGTTTCATG GCTGTTATTT	2415
	TATGGGAATT ATGGGCACTT GTATTGAGGT GGGGGCGGGA GCTTACATTC GGGGTTGTGA	2475
	GTTTGTGGGC TGTACCCTGG GAATCTGTTT TACTTCTAAG AGAGATATTA AGGTGAGGCA	2535
15	GTGCAACTTT GACAAATGCT TACTGGGTAT TACTTGTAAG GGGGACTATC GTCTTTCGGG	2595
	AAATGTGTGT TCTGAGACTT TCTGCTTTCG TCATTAGAG GGAGAGGGTT TGGTTAAAAA	2655
	CAACACAGTC AAGTCCCTTA GTCGCTGGAC CAGCGAGTCT GGCTTTTCCA TGATAACTTG	2715
	TGCAGACGGC AGGGTTACGC CTTTGGGTTT CCTCCACATT GTGGGCAACC GTTGTAGGCG	2775
	TTGGCCAACC ATGCAGGGGA ATGTGTTTAT CATGTCTAAA CTGTATCTGG GCAACAGAAT	2835
20	AGGGACTGTA GCCCTGCCCC AGTGTGCTTT CTACAAGTCC AGCATTGTGT TGGAGGAGAG	2895
	GGCGACAAAC AAGCTGGTCT TGGCTTGTGC TTTTGAGAAT AATGTACTGG TGTACAAAGT	2955
	GCTGAGACGG GAGAGTCCCT CAACCGTGAA AATGTGTGTT TGTGGGACTT CTCATTATGC	3015
	AAAGCCTTTG ACACTGGCAA TTATTTCTTC AGATATTCGG GCTAATCGAT ACATGTACAC	3075
	TGTGGAATCA ACAGAGTTCA CTCTGACGA GGATTAAGAG TGGGCGGGGC CAAGAGGGGT	3135
25	ATAAATAGGT GGGGAGGTTG AGGGGAGCCG TAGTTTCTGT TTTTCCAGA CTGGGGGGGA	3195
	CAACATGGCC GAGGAAGGGC GCATTATGT GCCTTATGTA ACTGCCCGCC TGCCCAAGTG	3255
	GTCGGGTTG GTGCAGGATA AGACGGGCTC GAACATGTTG GGGGGTGTGG TACTCCCTCC	3315
	TAATTCACAG GCGCACCGGA CGGAGACCGT GGGCACTGAG GCCACCAGAG ACAACCTGCA	3375
	CGCCGAGGGA GCGCGTCGTC CTGAGGATCA GACGCCCTAC ATGATCTTGG TGGAGGACTC	3435
30	TCTGGGAGGT TTGAAGAGGC GAATGGACTT GCTGGAAGAA TCTAATCAGC AGCTGCTGGC	3495
	AACTCTCAAC CGTCTCCGTA CAGGACTCGC TGCCTATGTG CAGGCTAACC TTGTGGGCGG	3555
	CCAAGTTAAC CCCTTTGTTT AAATAAAAT ACACTCATAC AGTTTATTAT GCTGTCAATA	3615
	AAATTCTTTA TTTTCTGT GATAATACCG TGTCCAGCGT GCTCTGTCAA TAAGGGTCCT	3675
	ATGCATCTG AGAAGGGCCT CATATACCCA TGGCATGAAT ATTAAGATAC ATGGGCATAA	3735
35	GGCCCTCAGA AGGGTTGAGG TAGAGCCACT GCAGACTTTC GTGGGAGGT AAGGTGTTGT	3795
	AAATAATCCA GTCATACTGA CTGTGCTGGG CGTGGAAGGA AAAGATGTCT TTTAGAAGAA	3855
	GGGTGATTGG CAAAGGGAGG CTCTTAGTGT AGGTATTGAT AAATCTGTTT AGTTGGGAGG	3915
	GATGCATTCG GGGGCTAATA AGGTGGAGTT TAGCCTGAAT CTTAAGGTTG GCAATGTTGC	3975
	CCCCTAGGTC TTTGCGAGGA TTCATGTTGT GCAGTACCAC AAAACAGAG TAGCCTGTGC	4035

-60-

ATTGGGGAA TTTATCATGA AGCTT

4060

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 211 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Lys Tyr Leu Val Leu Val Leu Asn Asp Gly Met Ser Arg Ile Glu
 1 5 10 15
 Lys Ala Leu Leu Cys Ser Asp Gly Glu Val Asp Leu Glu Cys His Glu
 20 25 30
 Val Leu Pro Pro Ser Pro Ala Pro Val Pro Ala Ser Val Ser Pro Val
 35 40 45
 Arg Ser Pro Pro Pro Leu Ser Pro Val Phe Pro Pro Ser Pro Pro Ala
 50 55 60
 Pro Leu Val Asn Pro Glu Ala Ser Ser Leu Leu Gln Gln Tyr Arg Arg
 65 70 75 80
 Glu Leu Leu Glu Arg Ser Leu Leu Arg Thr Ala Glu Gly Gln Gln Arg
 85 90 95
 Ala Val Cys Pro Cys Glu Arg Leu Pro Val Glu Glu Asp Glu Cys Leu
 100 105 110
 Asn Ala Val Asn Leu Leu Phe Pro Asp Pro Trp Leu Asn Ala Ala Glu
 115 120 125
 Asn Gly Gly Asp Ile Phe Lys Ser Pro Ala Met Ser Pro Glu Pro Trp
 130 135 140
 Ile Asp Leu Ser Ser Tyr Asp Ser Asp Val Glu Glu Val Thr Ser His
 145 150 155 160
 Phe Phe Leu Asp Cys Pro Glu Asp Pro Ser Arg Glu Cys Ser Ser Cys
 165 170 175
 Gly Phe His Gln Ala Gln Ser Gly Ile Pro Gly Ile Met Cys Ser Leu
 180 185 190
 Cys Tyr Met Arg Gln Thr Tyr His Cys Ile Tyr Ser Pro Val Ser Glu
 195 200 205
 Glu Glu Met
 210

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4060 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
 (B) LOCATION: 1476..1946

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CATCATCAAT AATCTACAGT ACACTGATGG CAGCGGTCCA ACTGCCAATC ATTTTGGCA 60
 CGTCATTAT GACGCAACGA CGGCGAGCGT GCGTGCTGA CGTAACTGTG GGGCGGAGCG 120

	CGTCGCGGAG GCGGCGGCGC TGGGCGGGGC TGAGGGCGGC GGGGCGGCGC GCGGGGCGCG	180
	GCGGCGGGGC GGGGCGAGGG GCGGAGTTCC GCACCCGCTA CGTCATTTTC AGACATTTT	240
	TAGCAAATTT GCGCCTTTTG CAAGCATTTT TCTCACATTT CAGGTATTTA GAGGGCGGAT	300
	TTTTGGTGTT CGTACTTCCG TGTCACATAG TTTACTGTCA ATCTTCATTA CGGCTTAGAC	360
5	AAATTTTCGG CGTCTTTTCC GGGTTTATGT CCCCGGTCAC CTTTATGACT GTGTGAAACA	420
	CACCTGCCCA TTGTTTACCC TTGGTCAGTT TTTTCGTCTC CTAGGGTGGG AACATCAAGA	480
	ACAAATTTGC CGAGTAATTG TGCACCTTTT TCCGCGTTAG GACTGCGTTT CACACGTAGA	540
	CAGACTTTTT CTCATTTTCT CACACTCCGT CGTCCGCTTC AGAGCTCTGC GTCTTCGCTG	600
	CCACCATGAA GTACCTGGTC CTCGTTCTCA ACGACGGCAT GAGTCGAATT GAAAAAGCTC	660
10	TCCTGTGCAG CGATGGTGAG GTGGATTTAG AGTGTCTATG GGTACTTCCC CTTTCTCCCG	720
	GCGCTGTCCC CGCTTCTGTG TCACCCGTGA GGAGTCTCC TCCTCTGTCT CCGGTGTTTC	780
	CTCCGTCTCC GCCAGCCCCG CTTGTGAATC CAGAGGCGAG TTCGCTGCTG CAGCAGTATC	840
	GGAGAGAGCT GTTAGAGAGG AGCCTGCTCC GAACGGCCGA AGGTCAGCAG CGTGCAGTGT	900
	GTCCATGTGA GCGGTTGCCC GTGGAAGAGG ATGAGTGTCT GAATGCCGTA AATTGCTGT	960
15	TTCTGATCC CTGGCTAAAT GCAGCTGAAA ATGGGGGTGA TATTTTAAAG TCTCCGGCTA	1020
	TGTCTCCAGA ACCGTGGATA GATTTGTCTA GCTACGATAG CGATGTAGAA GAGGTGACTA	1080
	GTCACTTTTT TCTGGATTGC CCTGAAGACC CCAGTCGGGA GTGTTTCTCT TGTGGGTTTC	1140
	ATCAGGCTCA AAGCGGAATT CCAGGCATTA TGTGCAGTTT GTGCTACATG CGCCAAACCT	1200
	ACCATTGCAT CTATAGTAAG TACATTCTGT AAAAGAACAT CTTGGTGATT TCTAGGTATT	1260
20	GTTTAGGGAT TAACTGGGTG GAGTGATCTT AATCCGGCAT AACCAAATAC ATGTTTTCAC	1320
	AGGTCCAGTT TCTGAAGAGG AAATGTGAGT CATGTTGACT TTGGCGCGCA AGAGGAAATG	1380
	TGAGTCATGT TGACTTTGGC GCGCCCTACG GTGACTTTAA AGCAATTTGA GGATCACTTT	1440
	TTTGTAGTC GCTATAAAGT AGTCACGGAG TCTTC ATG GAT CAC TTA AGC GTT	1493
	Met Asp His Leu Ser Val	
	1 5	
25	CTT TTG GAT TTG AAG CTG CTT CGC TCT ATC GTA GCG GGG GCT TCA AAT	1541
	Leu Leu Asp Leu Lys Leu Leu Arg Ser Ile Val Ala Gly Ala Ser Asn	
	10 15 20	
	CGC ACT GGA GTG TGG AAG AGG CGG CTG TGG CTG GGA CGC CTG ACT CAA	1589
	Arg Thr Gly Val Trp Lys Arg Arg Leu Trp Leu Gly Arg Leu Thr Gln	
	25 30 35	
	CTG GTC CAT GAT ACC TGC GTA GAG AAC GAG AGC ATA TTT CTC AAT TCT	1637
	Leu Val His Asp Thr Cys Val Glu Asn Glu Ser Ile Phe Leu Asn Ser	
30	40 45 50	
	CTG CCA GGG AAT GAA GCT TTT TTA AGG TTG CTT CGG AGC GGC TAT TTT	1685
	Leu Pro Gly Asn Glu Ala Phe Leu Arg Leu Leu Arg Ser Gly Tyr Phe	
	55 60 65 70	
	GAA GTG TTT GAC GTG TTT GTG GTG CCT GAG CTG CAT CTG GAC ACT CCG	1733
	Glu Val Phe Asp Val Phe Val Val Pro Glu Leu His Leu Asp Thr Pro	
	75 80 85	
35	GGT CGA GTG GTC GCC GCT CTT GCT CTG CTG GTG TTC ATC CTC AAC GAT	1781
	Gly Arg Val Val Ala Ala Leu Ala Leu Leu Val Phe Ile Leu Asn Asp	
	90 95 100	
	TTA GAC GCT AAT TCT GCT TCT TCA GGC TTT GAT TCA GGT TTT CTC GTG	1829
	Leu Asp Ala Asn Ser Ala Ser Ser Gly Phe Asp Ser Gly Phe Leu Val	
	105 110 115	
	GAC CGT CTC TGC GTG CCG CTA TGG CTG AAG GCC AGG GCG TTC AAG ATC	1877

	Asp	Arg	Leu	Cys	Val	Pro	Leu	Trp	Leu	Lys	Ala	Arg	Ala	Phe	Lys	Ile	
	120						125					130					
	ACC	CAG	AGC	TCC	AGG	AGC	ACT	TCG	CAG	CCT	TCC	TCG	TCG	CCC	GAC	AAG	1925
	Thr	Gln	Ser	Ser	Arg	Ser	Thr	Ser	Gln	Pro	Ser	Ser	Ser	Pro	Asp	Lys	
	135					140					145				150		
5	ACG	ACC	CAG	ACT	ACC	AGC	CAG	TAGACGGGGA	CAGCCCACCC	C6GGCTAGCC							1976
	Thr	Thr	Gln	Thr	Thr	Ser	Gln										
					155												
	TGGAGGAGGC	TGAACAGAGC	AGCACTCGTT	TCGAGCACAT	CAGTTACCGA	GACGTGGTGG											2036
	ATGACTTCAA	TAGATGCCAT	GATGTTTTTT	ATGAGAGGTA	CAGTTTTGAG	GACATAAAGA											2096
	GCTACGAGGC	TTTGCTTGAG	GACAATTTGG	AGCAGCTCAT	AGCTATGCAT	GCTAAAAATCA											2156
10	AGCTGCTGCC	CGGTGCGGAG	TATGAGTTGA	CTCAACCTTT	GAACATAACA	TCTYCGCCCT											2216
	ATGTGCTCGG	AAATGGGGCT	ACTATTAGGG	TAACAGGGGA	AGCCTCCCCG	GCTATTAGAG											2276
	TGGGGGCCAT	GGCCGTGGGT	CCGTGTGTAA	CAGGAATGAC	TGGGGTGACT	TTTGTGAATT											2336
	GTAGGTTTGA	GAGAGAGTCA	ACAATTAGGG	GGTCCCTGAT	ACGAGCTTCA	ACTCACGTGC											2396
	TGTTTCATGG	CTGTTATTTT	ATGGGAATTA	TGGGCACCTG	TATTGAGGTG	GGGGCGGGAG											2456
15	CTTACATTCG	GGGTTGTGAG	TTTGTGGGCT	GTTACCGGGG	AATCTGTTCT	ACTTCTAACA											2516
	GAGATATTAA	GGTGAGGCAG	TGCAACTTTG	ACAAATGCTT	ACTGGGTATT	ACTTGTAAGG											2576
	GGGACTATCG	TCTTTCGGGA	AATGTGTGTT	CTGAGACTTT	CTGCTTTGCT	CATTTAGAGG											2636
	GAGAGGGTTT	GGTTAAAAAC	AACACAGTCA	AGTCCCCTAG	TCGCTGGACC	AGCGAGTCTG											2696
	GCTTTTCCAT	GATAACTTGT	GCAGACGGCA	GGGTTACGCC	TTTGGGTTC	CTCCACATTG											2756
20	TGGGCAACCG	TTGTAGGCGT	TGGCCAACCA	TGCAGGGGAA	TGTGTTTATC	ATGTCTAAAC											2816
	TGTATCTGGG	CAACAGAATA	GGGACTGTAG	CCCTGCCCCA	GTGTGCTTTC	TACAAGTCCA											2876
	GCATTGTGTT	GGAGGAGAGG	GGCACAACCA	AGCTGGTCTT	GGCTTGTGCT	TTTGAGAATA											2936
	ATGTACTGGT	GTACAAAGTG	CTGAGACGGG	AGAGTCCCTC	AACCGTGAAA	ATGTGTGTTT											2996
	GTGGGACTTC	TCATTATGCA	AAGCCTTTGA	CACTGGCAAT	TATTTCTTCA	GATATTCGGG											3056
25	CTAATCGATA	CATGTACACT	GTGGACTCAA	CAGAGTTCAC	TTCTGACGAG	GATTAAGT											3116
	GGGCGGGGCC	AAGAGGGGTA	TAAATAGGTT	GGGAGGTTGA	GGGAGCCCGT	AGTTTCTGTT											3176
	TTTCCCAGAC	TGGGGGGGAC	AACATGGCCG	AGGAAGGGCG	CATTTATGTG	CCTTATGTAA											3236
	CTGCCCGCCT	GCCCAAGTGG	TCGGGTTCCG	TGCAGGATAA	GACGGGCTCG	AACATGTTGG											3296
	GGGGTGTGGT	ACTCCCTCCT	AATTACACAGG	CGCACCGGAC	GGAGACCGTG	GGCACTGAGG											3356
30	CCACCAGAGA	CAACCTGCAC	GCCGAGGGAG	CGCGTCGTCC	TGAGGATGAC	ACGCCCTACA											3416
	TGATCTTGGT	GGAGGACTCT	CTGGGAGGTT	TGAAGAGGCG	AATGGACTTG	CTGGAAGAAT											3476
	CTAATCAGCA	GCTGCTGGCA	ACTCTCAACC	GTCTCCGTAC	AGGACTCGCT	GCCTATGTGC											3536
	AGGCTAACCT	TGTGGGCGGC	CAAGTTAACC	CCTTTGTTTA	AATAAAAAATA	CACTCATACA											3596
	GTTTATTATG	CTGTCAATAA	AATTCTTTAT	TTTTCTGTG	ATAATACCGT	GTCCAGCGTG											3656
35	CTCTGTCAAT	AAGGGTCTTA	TGCATCCTGA	GAAGGGCCTC	ATATACCCAT	GGCATGAATA											3716
	TTAAGATACA	TGGGCATAAG	GCCCTCAGAA	GGGTTGAGGT	AGAGCCACTG	CAGACTTTTCG											3776
	TGGGGAGGTA	AGGTGTTGTA	AATAATCCAG	TCATACTGAC	TGTGCTGGGC	GTGGAAGGAA											3836
	AAGATGTCTT	TTAGAAGAAG	GGTGATTGGC	AAAGGAGGCG	TCTTAGTGTA	GGTATTGATA											3896
	AATCTGTTCA	GTTGGGAGGG	ATGCATTCCG	GGGCTAATAA	GGTGAGATT	AGCCTGAATC											3956

-63-

TTAAGGTTGG CAATGTTGCC CCCTAGGTCT TTGCGAGGAT TCATGTTGTG CAGTACCACA 4016
 AAAACAGAGT AGCCTGTGCA TTTGGGGAAT TTATCATGAA GCTT 4060

(2) INFORMATION FOR SEQ ID NO:4:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 157 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

10 Met Asp His Leu Ser Val Leu Leu Asp Leu Lys Leu Leu Arg Ser Ile
 1 5 10 15
 Val Ala Gly Ala Ser Asn Arg Thr Gly Val Trp Lys Arg Arg Leu Trp
 20 25 30
 Leu Gly Arg Leu Thr Gln Leu Val His Asp Thr Cys Val Glu Asn Glu
 35 40 45
 Ser Ile Phe Leu Asn Ser Leu Pro Gly Asn Glu Ala Phe Leu Arg Leu
 50 55 60
 15 Leu Arg Ser Gly Tyr Phe Glu Val Phe Asp Val Phe Val Val Pro Glu
 65 70 75 80
 Leu His Leu Asp Thr Pro Gly Arg Val Val Ala Ala Leu Ala Leu Leu
 85 90 95
 Val Phe Ile Leu Asn Asp Leu Asp Ala Asn Ser Ala Ser Ser Gly Phe
 100 105 110
 Asp Ser Gly Phe Leu Val Asp Arg Leu Cys Val Pro Leu Trp Leu Lys
 115 120 125
 20 Ala Arg Ala Phe Lys Ile Thr Gln Ser Ser Arg Ser Thr Ser Gln Pro
 130 135 140
 Ser Ser Ser Pro Asp Lys Thr Thr Gln Thr Thr Ser Gln
 145 150 155

(2) INFORMATION FOR SEQ ID NO:5:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4060 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- 30 (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 1850..3109

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CATCATCAAT AATCTACAGT ACACTGATGG CAGCGGTCCA ACTGCCAATC ATTTTGTCCA 60
 CGTCATTAT GACGCAACGA CGGCGAGCGT GCGGTGCTGA CGTAACTGTG GGGCGGAGCG 120
 35 CGTCGCGGAG GCGGCGGCGC TGGGCGGGGC TGAGGCGGCG GGGGCGGCG CGCGGGGCGG 180
 CGCGCGGGGC GGGGCGAGGG GCGGAGTTCC GCACCCGCTA CGTCATTTTC AGACATTTTT 240
 TAGCAAATTT GCGCCTTTTG CAAGCATTTT TCTCACATT CAGGTATTTA GAGGCGGAT 300
 TTTTGGTGTG CGTACTTCCG TGTACATAG TTCACTGTCA ATCTTCATTA CGGCTTAGAC 360
 AAATTTTCGG CGTCTTTTCC GGGTTTATGT CCCCGGTCAC CTTTATGACT GTGTGAAACA 420

-64-

	CACCTGCCCA TTGTTTACCC TTGGTCAGTT TTTTCGTCTC CTAGGGTGGG AACATCAAGA	480
	ACAAATTTCG CGAGTAATTG TGCACCTTTT TCCGCGTTAG GACTGCGTTT CACACGTAGA	540
	CAGACTTTTT CTCATTTTCT CACACTCCGT CGTCCGCTTC AGAGCTCTGC GTCTTCGCTG	600
	CCACCATGAA GTACCTGGTC CTCGTTCTCA ACGACGGCAT GAGTCGAATT GAAAAAGCTC	660
5	TCCTGTGCAG CGATGGTGAG GTGGATTAG AGTGTCATGA GGTACTTCCC CCTTCTCCCG	720
	CGCCTGTCCC CGCTTCTGTG TCACCCGTGA GGAGTCTCC TCCTCTGTCT CCGGTGTTTC	780
	CTCCGCTCC GCCAGCCCCG CTTGTGAATC CAGAGGCGAG TTCGCTGCTG CAGCAGTATC	840
	GGAGAGAGCT GTTAGAGAGG AGCCTGCTCC GAACGGCCGA AGGTCAGCAG CGTGCACTGT	900
	GTCCATGTGA GCGGTTGCCC GTGGAAGAGG ATGAGTGTCT GAATGCCGA AATTTGCTGT	960
10	TTCTGATCC CTGGCTAAAT GCAGCTGAAA ATGGGGGTGA TATTTTAAAG TCTCCGGCTA	1020
	TGTCTCCAGA ACCGTGGATA GATTTGTCTA GCTACGATAG CGATGTAGAA GAGGTGACTA	1080
	GTCACTTTTT TCTGGATTGC CCTGAAGACC CCAGTCGGGA GTGTTTCATCT TGTGGGTTTC	1140
	ATCAGGCTCA AAGCGGAATT CCAGGCATTA TGTGCAGTTT GTGCTACATG CGCCAAACCT	1200
	ACCATTCAT CTATAGTAAG TACATTCTGT AAAAGAACAT CTGTGTGATT TCTAGGTATT	1260
15	GTTTAGGGAT TAACTGGGTG GAGTGATCTT AATCCGGCAT AACCAATAC ATGTTTTCAC	1320
	AGGTCCAGTT TCTGAAGAGG AAATGTGAGT CATGTTGACT TTGGCGCGCA AGAGGAAATG	1380
	TGAGTCATGT TGACTTTGGC GCGCCCTACG GTGACTTTAA AGCAATTTGA GGATCACTTT	1440
	TTTGTTAGTC GCTATAAAGT AGTCACGGAG TCTTCATGGA TCACTTAAGC GTTCTTTTGG	1500
	ATTTGAAGCT GCTTCGCTCT ATCGTAGCGG GGGCTTCAAA TCGCACTGGA GTGTGGAAGA	1560
20	GGCGGCTGTG GCTGGGACGC CTGACTCAAC TGGTCCATGA TACCTGGCTA GAGAACGAGA	1620
	GCATATTCTT CAATTCTCTG CCAGGGAATG AAGCTTTTTT AAGTTTGCTT CGGAGCGGCT	1680
	ATTTTGAAGT GTTTGACGTG TTTGTGGTGC CTGAGCTGCA TCTGGACACT CCGGGTCGAG	1740
	TGGTCGCCGC TCTTGCTCTG CTGGTGTTC TCCTCAACGA TTTAGACGCT AATTCTGCTT	1800
	CTTCAGGCTT TGATTCAGGT TTTCTCGTGG ACCGTCTCTG CGTGCCGCT ATG GCT	1855
25		Met Ala 1
	GAA GGC CAG GGC GTT CAA GAT CAC CCA GAG CTC CAG GAG CAC TTC GCA	1903
	Glu Gly Gln Gly Val Gln Asp His Pro Glu Leu Gln Glu His Phe Ala	
	5 10 15	
	GCC TTC CTC GTC GCC CGA CAA GAC GAC CCA GAC TAC CAG CCA GTA GAC	1951
	Ala Phe Leu Val Ala Arg Gln Asp Asp Pro Asp Tyr Gln Pro Val Asp	
	20 25 30	
30	GGG GAC AGC CCA CCC CGG GCT AGC CTG GAG GAG GCT GAA CAG AGC AGC	1999
	Gly Asp Ser Pro Pro Arg Ala Ser Leu Glu Glu Ala Glu Gln Ser Ser	
	35 40 45 50	
	ACT CGT TTC GAG CAC ATC AGT TAC CGA GAC GTG GTG GAT GAC TTC AAT	2047
	Thr Arg Phe Glu His Ile Ser Tyr Arg Asp Val Val Asp Asp Phe Asn	
	55 60 65	
	AGA TGC CAT GAT GTT TTT TAT GAG AGG TAC AGT TTT GAG GAC ATA AAG	2095
	Arg Cys His Asp Val Phe Tyr Glu Arg Tyr Ser Phe Glu Asp Ile Lys	
35	70 75 80	
	AGC TAC GAG GCT TTG CCT GAG GAC AAT TTG GAG CAG CTC ATA GCT ATG	2143
	Ser Tyr Glu Ala Leu Pro Glu Asp Asn Leu Glu Gln Leu Ile Ala Met	
	85 90 95	
	CAT GCT AAA ATC AAG CTG CTG CCC GGT CGG GAG TAT GAG TTG ACT CAA	2191
	His Ala Lys Ile Lys Leu Leu Pro Gly Arg Glu Tyr Glu Leu Thr Gln	
	100 105 110	

-65-

	CCT TTG AAC ATA ACA TCT TGC GCC TAT GTG CTC GGA AAT GGG GCT ACT Pro Leu Asn Ile Thr Ser Cys Ala Tyr Val Leu Gly Asn Gly Ala Thr 115 120 125 130	2239
	ATT AGG GTA ACA GGG GAA GCC TCC CCG GCT ATT AGA GTG GGG GCC ATG Ile Arg Val Thr Gly Glu Ala Ser Pro Ala Ile Arg Val Gly Ala Met 135 140 145	2287
5	GCC GTG GGT CCG TGT GTA ACA GGA ATG ACT GGG GTG ACT TTT GTG AAT Ala Val Gly Pro Cys Val Thr Gly Met Thr Gly Val Thr Phe Val Asn 150 155 160	2335
	TGT AGG TTT GAG AGA GAG TCA ACA ATT AGG GGG TCC CTG ATA CGA GCT Cys Arg Phe Glu Arg Glu Ser Thr Ile Arg Gly Ser Leu Ile Arg Ala 165 170 175	2383
10	TCA ACT CAC GTG CTG TTT CAT GGC TGT TAT TTT ATG GGA ATT ATG GGC Ser Thr His Val Leu Phe His Gly Cys Tyr Phe Met Gly Ile Met Gly 180 185 190	2431
	ACT TGT ATT GAG GTG GGG GCG GGA GCT TAC ATT CGG GGT TGT GAG TTT Thr Cys Ile Glu Val Gly Ala Gly Ala Tyr Ile Arg Gly Cys Glu Phe 195 200 205 210	2479
	GTG GGC TGT TAC CGG GGA ATC TGT TCT ACT TCT AAC AGA GAT ATT AAG Val Gly Cys Tyr Arg Gly Ile Cys Ser Thr Ser Asn Arg Asp Ile Lys 215 220 225	2527
15	GTG AGG CAG TGC AAC TTT GAC AAA TGC TTA CTG GGT ATT ACT TGT AAG Val Arg Gln Cys Asn Phe Asp Lys Cys Leu Leu Gly Ile Thr Cys Lys 230 235 240	2575
	GGG GAC TAT CGT CTT TCG GGA AAT GTG TGT TCT GAG ACT TTC TGC TTT Gly Asp Tyr Arg Leu Ser Gly Asn Val Cys Ser Glu Thr Phe Cys Phe 245 250 255	2623
20	GCT CAT TTA GAG GGA GAG GGT TTG GTT AAA AAC AAC ACA GTC AAG TCC Ala His Leu Glu Gly Glu Gly Leu Val Lys Asn Asn Thr Val Lys Ser 260 265 270	2671
	CCT AGT CGC TGG ACC AGC GAG TCT GGC TTT TCC ATG ATA ACT TGT GCA Pro Ser Arg Trp Thr Ser Glu Ser Gly Phe Ser Met Ile Thr Cys Ala 275 280 285 290	2719
	GAC GGC AGG GTT ACG CCT TTG GGT TCC CTC CAC ATT GTG GGC AAC CGT Asp Gly Arg Val Thr Pro Leu Gly Ser Leu His Ile Val Gly Asn Arg 295 300 305	2767
25	TGT AGG CGT TGG CCA ACC ATG CAG GGG AAT GTG TTT ATC ATG TCT AAA Cys Arg Arg Trp Pro Thr Met Gln Gly Asn Val Phe Ile Met Ser Lys 310 315 320	2815
	CTG TAT CTG GGC AAC AGA ATA GGG ACT GTA GCC CTG CCC CAG TGT GCT Leu Tyr Leu Gly Asn Arg Ile Gly Thr Val Ala Leu Pro Gln Cys Ala 325 330 335	2863
30	TTC TAC AAG TCC AGC ATT TGT TTG GAG GAG AGG GCG ACA AAC AAG CTG Phe Tyr Lys Ser Ser Ile Cys Leu Glu Glu Arg Ala Thr Asn Lys Leu 340 345 350	2911
	GTG TTG GCT TGT GCT TTT GAG AAT AAT GTA CTG GTG TAC AAA GTG CTG Val Leu Ala Cys Ala Phe Glu Asn Asn Val Leu Val Tyr Lys Val Leu 355 360 365 370	2959
	AGA CGG GAG AGT CCC TCA ACC GTG AAA ATG TGT GTT TGT GGG ACT TCT Arg Arg Glu Ser Pro Ser Thr Val Lys Met Cys Val Cys Gly Thr Ser 375 380 385	3007
35	CAT TAT GCA AAG CCT TTG ACA CTG GCA ATT ATT TCT TCA GAT ATT CGG His Tyr Ala Lys Pro Leu Thr Leu Ala Ile Ile Ser Ser Asp Ile Arg 390 395 400	3055
	GCT AAT CGA TAC ATG TAC ACT GTG GAC TCA ACA GAG TTC ACT TCT GAC Ala Asn Arg Tyr Met Tyr Thr Val Asp Ser Thr Glu Phe Thr Ser Asp 405 410 415	3103
	GAG GAT TAAAGTGGG CGGGGCAAG AGGGGTATAA ATAGGTGGGG AGGTTGAGGG	3159

	Glu Asp	
	420	
	GAGCCGTAGT TTCTGTTTT CCCAGACTGG GGGGGACAAC ATGGCCGAGG AAGGGCGCAT	3219
	TTATGTGCCT TATGTAACG CCCGCCTGCC CAAGTGGTCG GGTTCGGTGC AGGATAAGAC	3279
5	GGGCTCGAAC ATGTTGGGGG GTGTGGTACT CCCTCCTAAT TCACAGGCGC ACCGGACGGA	3339
	GACCGTGGGC ACTGAGGCCA CCAGAGACAA CCTGCACGCC GAGGGAGCGC GTCGTCCTGA	3399
	GGATCAGACG CCCTACATGA TCTTGGTGA GGA CTCTCTG GGAGGTTTGA AGAGGCGAAT	3459
	GGACTTGCTG GAAGAATCTA ATCAGCAGCT GCTGGCAACT CTCAACCGTC TCCGTACAGG	3519
	ACTCGCTGCC TATGTGCAGG CTAACCTTGT GGGCGGCCAA GTTAACCCCT TTGTTTAAAT	3579
10	AAAAATACAC TCATACAGTT TATTATGCTG TCAATAAAAT TCTTTATTTT TCCTGTGATA	3639
	ATACCGTGTC CAGCGTGCTC TGTCATAAG GGTCTATGC ATCCTGAGAA GGGCCTCATA	3699
	TACCCATGCC ATGAATATTA AGATACATGG GCATAAGGCC CTCAGAAGGG TTGAGGTAGA	3759
	GCCACTGCAG ACTTTCGTGG GGAGGTAAGG TGTGTAAAT AATCCAGTCA TACTGACTGT	3819
	GCTGGGCGTG GAAGGAAAAG ATGTCCTTTA GAAGAAGGGT GATTGGCAAA GGGAGGCTCT	3879
15	TAGTGTAGGT ATTGATAAAT CTGTTCAAGT GGGAGGGATG CATTGCGGGG CTAATAAGGT	3939
	GGAGTTTAGC CTGAATCTTA AGGTTGGCAA TGTGCCCCC TAGGTCTTTG CGAGGATTCA	3999
	TGTTGTGCAG TACCACAAAA ACAGAGTAGC CTGTGCATTT GGGGAATTTA TCATGAAGCT	4059
	T	4060

(2) INFORMATION FOR SEQ ID NO:6:

20	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 420 amino acids
	(B) TYPE: amino acid
	(D) TOPOLOGY: linear
	(ii) MOLECULE TYPE: protein
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
25	Met Ala Glu Gly Gln Gly Val Gln Asp His Pro Glu Leu Gln Glu His 1 5 10 15
	Phe Ala Ala Phe Leu Val Ala Arg Gln Asp Asp Pro Asp Tyr Gln Pro 20 25 30
	Val Asp Gly Asp Ser Pro Pro Arg Ala Ser Leu Glu Glu Ala Glu Gln 35 40 45
30	Ser Ser Thr Arg Phe Glu His Ile Ser Tyr Arg Asp Val Val Asp Asp 50 55 60
	Phe Asn Arg Cys His Asp Val Phe Tyr Glu Arg Tyr Ser Phe Glu Asp 65 70 75 80
	Ile Lys Ser Tyr Glu Ala Leu Pro Glu Asp Asn Leu Glu Gln Leu Ile 85 90 95
	Ala Met His Ala Lys Ile Lys Leu Leu Pro Gly Arg Glu Tyr Glu Leu 100 105 110
35	Thr Gln Pro Leu Asn Ile Thr Ser Cys Ala Tyr Val Leu Gly Asn Gly 115 120 125
	Ala Thr Ile Arg Val Thr Gly Glu Ala Ser Pro Ala Ile Arg Val Gly 130 135 140
	Ala Met Ala Val Gly Pro Cys Val Thr Gly Met Thr Gly Val Thr Phe 145 150 155 160

-67-

Val Asn Cys Arg Phe Glu Arg Glu Ser Thr Ile Arg Gly Ser Leu Ile
 165 170 175
 Arg Ala Ser Thr His Val Leu Phe His Gly Cys Tyr Phe Met Gly Ile
 180 185 190
 Met Gly Thr Cys Ile Glu Val Gly Ala Gly Ala Tyr Ile Arg Gly Cys
 195 200 205
 5 Glu Phe Val Gly Cys Tyr Arg Gly Ile Cys Ser Thr Ser Asn Arg Asp
 210 215 220
 Ile Lys Val Arg Gln Cys Asn Phe Asp Lys Cys Leu Leu Gly Ile Thr
 225 230 235 240
 Cys Lys Gly Asp Tyr Arg Leu Ser Gly Asn Val Cys Ser Glu Thr Phe
 245 250 255
 10 Cys Phe Ala His Leu Glu Gly Glu Gly Leu Val Lys Asn Asn Thr Val
 260 265 270
 Lys Ser Pro Ser Arg Trp Thr Ser Glu Ser Gly Phe Ser Met Ile Thr
 275 280 285
 Cys Ala Asp Gly Arg Val Thr Pro Leu Gly Ser Leu His Ile Val Gly
 290 295 300
 15 Asn Arg Cys Arg Arg Trp Pro Thr Met Gln Gly Asn Val Phe Ile Met
 305 310 315 320
 Ser Lys Leu Tyr Leu Gly Asn Arg Ile Gly Thr Val Ala Leu Pro Gln
 325 330 335
 Cys Ala Phe Tyr Lys Ser Ser Ile Cys Leu Glu Glu Arg Ala Thr Asn
 340 345 350
 Lys Leu Val Leu Ala Cys Ala Phe Glu Asn Asn Val Leu Val Tyr Lys
 355 360 365
 20 Val Leu Arg Arg Glu Ser Pro Ser Thr Val Lys Met Cys Val Cys Gly
 370 375 380
 Thr Ser His Tyr Ala Lys Pro Leu Thr Leu Ala Ile Ile Ser Ser Asp
 385 390 395 400
 Ile Arg Ala Asn Arg Tyr Met Tyr Thr Val Asp Ser Thr Glu Phe Thr
 405 410 415
 25 Ser Asp Glu Asp
 420

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4060 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
 (B) LOCATION: 3200..3574

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CATCATCAAT AATCTACAGT ACACTGATGG CAGCGGTCCA ACTGCCAATC ATTTTGGCCA	60
CGTCATTAT GACGCAACGA CGGCGAGCGT GCGTGCTGA CGTAACTGTG GGGCGGAGCG	120
CGTCGCGGAG GCGGCGGCGC TGGGCGGGGC TGAGGCGGCG GGGGCGGCGC CGCGGGGCGG	180
CGCGCGGGGC GGGGCGAGGG GCGGAGTTCC GCACCCGCTA CGTCATTTTC AGACATTTT	240

	TAGCAAATTT GCGCCTTTTG CAAGCATT TTCTACATTT CAGGTATTTA GAGGGCGGAT	300
	TTTTGGTGT CGTACTCCG TGTCACATAG TTCACTGTCA ATCTTCATTA CGGCTTAGAC	360
	AAATTTTCGG CGTCTTTTCC GGGTTATGT CCCC GGTCAC CTTTATGACT GTGTGAAACA	420
	CACCTGCCCC TTGTTTACCC TTGGTCAGTT TTTTCGTCTC CTAGGGTGGG AACATCAAGA	480
5	ACAAATTTGC CBAGTAATTG TGCACCTTTT TCCGCGTTAG GACTGCGTTT CACACGTAGA	540
	CAGACTTTTT CTCATTTTCT CACACTCCGT CGTCCGCTTC AGAGCTCTGC GTCTTCGCTG	600
	CCACCATGAA GTACCTGGTC CTCGTTCTCA ACGACGGCAT GAGTCGAATT GAAAAAGCTC	660
	TCCTGTGCAG CGATGGTGAG GTGGATTAG AGTGTCATGA GGTACTTCCC CTTTCTCCCG	720
	CGCCTGTCCC CGCTTCTGTG TCACCCGTGA GGAGTCTCC TCCTCTGTCT CCGGTGTTTC	780
10	CTCCGTCTCC GCCAGCCCCG CTTGTGAATC CAGAGGCGAG TTCGTGCTG CAGCAGTATC	840
	GGAGAGAGCT GTTAGAGAGG AGCCTGCTCC GAACGGCCGA AGGTGAGCAG CGTGCACTGT	900
	GTCCATGTGA GCGGTTGCCC GTGGAAGAGG ATGAGTGTCT GAATGCCGTA AATTGCTGT	960
	TTCTGATCC CTGGCTAAAT GCAGCTGAAA ATGGGGGTGA TATTTTAAAG TCTCCGGCTA	1020
	TGTCTCCAGA ACCGTGGATA GATTGTCTA GCTACGATAG CGATGTAGAA GAGGTGACTA	1080
15	GTCACTTTTT TCTGGATTGC CCTGAAGACC CCAGTCGGGA GTGTTTCTCT TGTGGGTTTC	1140
	ATCAGGCTCA AAGCGGAATT CCAGGCATTA TGTGCAGTTT GTGCTACATG CGCCAAACCT	1200
	ACCATTGCAT CTATAGTAAG TACATTCTGT AAAAGAACAT CTGGGTGATT TCTAGGTATT	1260
	GTTTAGGGAT TAACTGGGTG GAGTGATCTT AATCCGGCAT AACCAAATAC ATGTTTTCAC	1320
	AGGTCCAGTT TCTGAAGAGG AAATGTGAGT CATGTTGACT TTGGCGCGCA AGAGGAAATG	1380
20	TGAGTCATGT TGACTTTGGC GCGCCCTACG GTGACTTTAA AGCAATTTGA GGATCACTTT	1440
	TTTGTAGTC GCTATAAAGT AGTCACGGAG TCTTCATGGA TCACTTAAGC GTTCTTTTGG	1500
	ATTTGAAGCT GCTTCGCTCT ATCGTAGCGG GGGCTTCAAA TCGCACTGGA GTGTGGAAGA	1560
	GGCGGCTGTG GCTGGGACGC CTGACTCAAC TGGTCCATGA TACCTGCGTA GAGAACGAGA	1620
	GCATATTTCT CAATTCTCTG CCAGGGAATG AAGCTTTTTT AAGGTTGCTT CGGAGCGGCT	1680
25	ATTTTGAAGT GTTTGACGTG TTTGTGGTGC CTGAGCTGCA TCTGGACACT CCGGGTCGAG	1740
	TGGTCGCGCG TCTTGCTCTG CTGGTGTCA TCCTCAACGA TTTAGACGCT AATTCTGCTT	1800
	CTTCAGGCTT TGATTCAGGT TTTCTCGTG ACCGTCTCTG CGTGCCGCTA TGGCTGAAGG	1860
	CCAGGGCGTT CAAGATCACC CAGAGCTCCA GGAGCACTTC GCAGCCTTCC TCCTCGCCCG	1920
	ACAAGACGAC CCAGACTACC AGCCAGTAGA CGGGACAGC CCACCCCGGG CTAGCCTGGA	1980
30	GGAGGCTGAA CAGAGCAGCA CTCGTTTCGA GCACATCAGT TACCGAGACG TGGTGATGA	2040
	CTTCAATAGA TGCCATGATG TTTTATGA GAGGTACAGT TTTGAGGACA TAAAGAGCTA	2100
	CGAGGCTTTG CCTGAGGACA ATTTGGAGCA GTCATAGCT ATGCATGCTA AAATCAAGCT	2160
	GCTGCCCGGT CGGGAGTATG AGTTGACTCA ACCTTTGAAC ATAACATCTT GCGCCTATGT	2220
	GCTCGGAAAT GGGGCTACTA TTAGGGTAAC AGGGGAAGCC TCCCCGGCTA TTAGAGTGGG	2280
35	GGCCATGGCC GTGGGTCCGT GTGTAACAGG AATGACTGGG GTGACTTTTG TGAATTGTAG	2340
	GTTTGAGAGA GAGTCAACAA TTAGGGGGTC CCTGATACGA GCTTCAACTC ACGTGCTGTT	2400
	TCATGGCTGT TATTTTATGG GAATTATGGG CACTTGTATT GAGGTGGGGG CGGGAGCTTA	2460
	CATTGCGGGT TGTGAGTTTG TGGGCTGTTA CCGGGGAATC TGTCTACTT CTAACAGAGA	2520
	TATTAAGGTG AGGCAGTGCA ACTTTGACAA ATGCTTACTG GGTATTACTT GTAAGGGGGA	2580

-69-

	CTATCGTCTT TCGGGAAATG TGTGTTCTGA GACTTTCTGC TTTGCTCATT TAGAGGGAGA	2640
	GGGTTTGGTT AAAACAACA CAGTCAAGTC CCCTAGTCGC TGGACCAGCG AGTCTGGCTT	2700
	TTCCATGATA ACTTGTGCAG ACGGCAGGGT TACGCCCTTG GGTCCCTCC ACATTGTGGG	2760
	CAACCGTTGT AGGCGTTGGC CAACCATGCA GGGGAATGTG TTTATCATGT CTAAACTGTA	2820
5	TCTGGGCAAC AGAATAGGGA CTGTAGCCCT GCCCAGTGT GCTTCTACA AGTCCAGCAT	2880
	TTGTTTGGAG GAGAGGGCGA CAAACAAGCT GGTCTTGGCT TGTGCTTTG AGAATAATGT	2940
	ACTGGTGTAC AAAGTGTGA GACGGGAGAG TCCCTCAACC GTGAAAATGT GTGTTTGTGG	3000
	GACTTCTCAT TATGCAAAGC CTTTGACACT GGCAATTATT TCTTCAGATA TTCGGGCTAA	3060
	TCGATACATG TACACTGTGG ACTCAACAGA GTTCACTTCT GACGAGGATT AAAAGTGGGC	3120
10	GGGGCCAAGA GGGGTATAAA TAGGTGGGGA GGTGAGGGG AGCCGTAGTT TCTGTTTTTC	3180
	CCGAGCTGGG GGGGACAAC ATG GCC GAG GAA GGG CGC ATT TAT GTG CCT TAT	3232
	Met Ala Glu Glu Gly Arg Ile Tyr Val Pro Tyr	
	1 5 10	
	GTA ACT GCC CGC CTG CCC AAG TGG TCG GGT TCG GTG CAG GAT AAG ACG	3280
	Val Thr Ala Arg Leu Pro Lys Trp Ser Gly Ser Val Gln Asp Lys Thr	
	15 20 25	
15	GGC TCG AAC ATG TTG GGG GGT GTG GTA CTC CCT CCT AAT TCA CAG GCG	3328
	Gly Ser Asn Met Leu Gly Gly Val Val Leu Pro Pro Asn Ser Gln Ala	
	30 35 40	
	CAC CGG ACG GAG ACC GTG GGC ACT GAG GCC ACC AGA GAC AAC CTG CAC	3376
	His Arg Thr Glu Thr Val Gly Thr Glu Ala Thr Arg Asp Asn Leu His	
	45 50 55	
	GCC GAG GGA GCG CGT CGT CCT GAG GAT CAG ACG CCC TAC ATG ATC TTG	3424
	Ala Glu Gly Ala Arg Arg Pro Glu Asp Gln Thr Pro Tyr Met Ile Leu	
20	60 65 70 75	
	GTG GAG GAC TCT CTG GGA GGT TTG AAG AGG CGA ATG GAC TTG CTG GAA	3472
	Val Glu Asp Ser Leu Gly Gly Leu Lys Arg Arg Met Asp Leu Leu Glu	
	80 85 90	
	GAA TCT AAT CAG CAG CTG CTG GCA ACT CTC AAC CGT CTC CGT ACA GGA	3520
	Glu Ser Asn Gln Gln Leu Leu Ala Thr Leu Asn Arg Leu Arg Thr Gly	
	95 100 105	
25	CTC GCT GCC TAT GTG CAG GCT AAC CTT GTG GGC GGC CAA GTT AAC CCC	3568
	Leu Ala Ala Tyr Val Gln Ala Asn Leu Val Gly Gly Gln Val Asn Pro	
	110 115 120	
	TTT GTT TAAATAAAAA TACACTCATA CAGTTTATTA TGCTGTCAAT AAAATTCTTT	3624
	Phe Val	
	125	
	ATTTTTCCTG TGATAATACC GTGTCCAGCG TGCTCTGTCA ATAAGGGTCC TATGCATCCT	3684
30	GAGAAGGGCC TCATATACCC ATGGCATGAA TATTAAGATA CATGGGCATA AGGCCCTCAG	3744
	AAGGGTTGAG GTAGAGCCAC TGCAGACTTT CGTGGGGAGG TAAGGTGTTG TAAATAATCC	3804
	AGTCATACTG ACTGTGCTGG GCGTGGAAGG AAAAGATGTC TTTTAGAAGA AGGGTGATTG	3864
	GCAAAGGGAG GCTCTTAGTG TAGGTATTGA TAAATCTGTT CAGTTGGGAG GGATGCATTG	3924
	GGGGGCTAAT AAGGTGGAGT TTAGCCTGAA TCTTAAGGTT GGCAATGTTG CCCCCTAGGT	3984
35	CTTTGCCAGG ATTCATGTTG TGCAGTACCA CAAAAACAGA GTAGCCTGTG CATTGGGGGA	4044
	ATTTATCATG AAGCTT	4060

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 125 amino acids

-70-

(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

5 Met Ala Glu Glu Gly Arg Ile Tyr Val Pro Tyr Val Thr Ala Arg Leu
1 5 10 15
Pro Lys Trp Ser Gly Ser Val Gln Asp Lys Thr Gly Ser Asn Met Leu
20 25 30
Gly Gly Val Val Leu Pro Pro Asn Ser Gln Ala His Arg Thr Glu Thr
35 40 45
Val Gly Thr Glu Ala Thr Arg Asp Asn Leu His Ala Glu Gly Ala Arg
50 55 60
10 Arg Pro Glu Asp Gln Thr Pro Tyr Met Ile Leu Val Glu Asp Ser Leu
65 70 75 80
Gly Gly Leu Lys Arg Arg Met Asp Leu Leu Glu Glu Ser Asn Gln Gln
85 90 95
Leu Leu Ala Thr Leu Asn Arg Leu Arg Thr Gly Leu Ala Ala Tyr Val
100 105 110
15 Gln Ala Asn Leu Val Gly Gly Gln Val Asn Pro Phe Val
115 120 125

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 54 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Glu Glu Phe Val Leu Asp Tyr Val Glu His Pro Gly His Gly Cys Arg
1 5 10 15
25 Ser Cys His Tyr His Arg Arg Asn Thr Gly Asp Pro Asp Ile Met Cys
20 25 30
Ser Leu Cys Tyr Met Arg Thr Cys Gly Met Phe Val Tyr Ser Pro Val
35 40 45
Ser Glu Pro Glu Pro Glu
50

(2) INFORMATION FOR SEQ ID NO:10:

30

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Ile Asp Leu Thr Cys His Glu Ala Gly Phe Pro Pro Ser
1 5 10

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

-71-

(A) LENGTH: 19 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Leu Asp Phe Ser Thr Pro Gly Arg Ala Ala Ala Val Ala Phe Leu
 1 5 10 15
 Ser Phe Ile

10

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 7 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Gln Ser Ser Asn Ser Thr Ser
 1 5

(2) INFORMATION FOR SEQ ID NO:13:

20

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 347 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Gln Lys Tyr Ser Ile Glu Gln Leu Thr Thr Tyr Trp Leu Gln Pro Gly
 1 5 10 15

Asp Asp Phe Glu Glu Ala Ile Arg Val Tyr Ala Lys Val Ala Leu Arg
 20 25 30

Pro Asp Cys Lys Tyr Lys Ile Ser Lys Leu Val Asn Ile Arg Asn Cys
 35 40 45

30

Cys Tyr Ile Ser Gly Asn Gly Ala Glu Val Glu Ile Asp Thr Glu Asp
 50 55 60

Arg Val Ala Phe Arg Cys Ser Met Ile Asn Met Trp Pro Gly Val Leu
 65 70 75 80

Gly Met Asp Gly Val Val Ile Met Asn Val Arg Phe Thr Gly Pro Asn
 85 90 95

Phe Ser Gly Thr Val Phe Leu Ala Asn Thr Asn Leu Ile Leu His Gly
 100 105 110

35

Val Ser Phe Tyr Gly Phe Asn Asn Thr Cys Val Glu Ala Trp Thr Asp
 115 120 125

Val Arg Val Arg Gly Cys Ala Phe Tyr Cys Cys Trp Lys Gly Val Val
 130 135 140

Cys Arg Pro Lys Ser Arg Ala Ser Ile Lys Lys Cys Leu Phe Glu Arg
 145 150 155 160

-72-

Cys Thr Leu Gly Ile Leu Ser Glu Gly Asn Ser Arg Val Arg His Asn
 165 170 175
 Val Ala Ser Asp Cys Gly Cys Phe Met Leu Val Lys Ser Val Ala Val
 180 185 190
 Ile Lys His Asn Met Val Cys Gly Asn Cys Glu Asp Arg Ala Ser Gln
 195 200 205
 Met Leu Thr Cys Ser Asp Gly Asn Cys His Leu Leu Lys Thr Ile His
 210 215 220
 Val Ala Ser His Ser Arg Lys Ala Trp Pro Val Phe Glu His Asn Ile
 225 230 235 240
 Leu His Arg Cys Ser Leu His Leu Gly Asn Arg Arg Gly Val Phe Leu
 245 250 255
 Pro Tyr Gln Cys Asn Leu Ser His Thr Lys Ile Leu Leu Glu Pro Glu
 260 265 270
 Ser Met Ser Lys Val Asn Leu Asn Gly Val Phe Asp Met Thr Met Lys
 275 280 285
 Ile Trp Lys Val Leu Arg Tyr Asp Glu Thr Arg Thr Arg Cys Arg Pro
 290 295 300
 Cys Glu Cys Gly Gly Lys His Ile Arg Asn Gln Pro Val Met Leu Asp
 305 310 315 320
 Val Thr Glu Glu Leu Arg Pro Asp His Leu Val Leu Ala Cys His Arg
 325 330 335
 Ala Glu Phe Gly Ser Ser Asp Glu Asp Thr Asp
 340 345

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 140 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:
 Met Ser Thr Asn Ser Phe Asp Gly Ser Ile Val Ser Ser Tyr Leu Thr
 1 5 10 15
 Thr Arg Met Pro Pro Trp Ala Gly Val Arg Gln Asn Val Met Gly Ser
 20 25 30
 Ser Ile Asp Gly Arg Pro Val Leu Pro Ala Asn Ser Thr Thr Leu Thr
 35 40 45
 Tyr Glu Thr Val Ser Gly Thr Pro Leu Glu Thr Ala Ala Ser Ala Ala
 50 55 60
 Ala Ser Ala Ala Ala Thr Ala Arg Gly Ile Val Thr Asp Phe Ala
 65 70 75 80
 Phe Leu Ser Pro Leu Ala Ser Ser Ala Ala Ser Arg Ser Ser Ala Arg
 85 90 95
 Asp Asp Lys Leu Thr Ala Leu Leu Ala Gln Leu Asp Ser Leu Thr Arg
 100 105 110
 Glu Leu Asn Val Val Ser Gln Gln Leu Leu Asp Leu Arg Gln Gln Val
 115 120 125
 Ser Ala Leu Lys Ala Ser Ser Pro Pro Asn Ala Val
 130 135 140

-73-

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5100 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
 (B) LOCATION: 2..418

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

10	C CTC ATC AAA CAA CCC GTG GTG GGC ACC ACC CAC GTG GAA ATG CCT Leu Ile Lys Gln Pro Val Val Gly Thr Thr His Val Glu Met Pro	46
	1 5 10 15	
	CGC AAC GAA GTC CTA GAA CAA CAT CTG ACC TCA CAT GGC GCT CAA ATC Arg Asn Glu Val Leu Glu Gln His Leu Thr Ser His Gly Ala Gln Ile	94
	20 25 30	
	GCG GGC GGA GGC GCT GCG GGC GAT TAC TTT AAA AGC CCC ACT TCA GCT Ala Gly Gly Gly Ala Ala Gly Asp Tyr Phe Lys Ser Pro Thr Ser Ala	142
15	35 40 45	
	CGA ACC CTT ATC CCG CTC ACC GCC TCC TGC TTA AGA CCA GAT GGA GTC Arg Thr Leu Ile Pro Leu Thr Ala Ser Cys Leu Arg Pro Asp Gly Val	190
	50 55 60	
	TTT CAA CTA GGA GGA GGC TCG CGT TCA TCT TTC AAC CCC CTG CAA ACA Phe Gln Leu Gly Gly Gly Ser Arg Ser Ser Phe Asn Pro Leu Gln Thr	238
	65 70 75	
20	GAT TTT GCC TTC CAC GCC CTG CCC TCC AGA CCG CGC CAC GGG GGC ATA Asp Phe Ala Phe His Ala Leu Pro Ser Arg Pro Arg His Gly Gly Ile	286
	80 85 90 95	
	GGA TCC AGG CAG TTT GTA GAG GAA TTT GTG CCC GCC GTC TAC CTC AAC Gly Ser Arg Gln Phe Val Glu Glu Phe Val Pro Ala Val Tyr Leu Asn	334
	100 105 110	
	CCC TAC TCG GGA CCG CCG GAC TCT TAT CCG GAC CAG TTT ATA CGC CAC Pro Tyr Ser Gly Pro Pro Asp Ser Tyr Pro Asp Gln Phe Ile Arg His	382
25	115 120 125	
	TAC AAC GTG TAC AGC AAC TCT GTG AGC GGT TAT AGC TGAGATTGA Tyr Asn Val Tyr Ser Asn Ser Val Ser Gly Tyr Ser	428
	130 135	
	AGACTCTCCT ATCTGTCTCT GTGCTGCTTT TCCGCTTCAA GCCCCACAAG CATGAAGGGG	488
	TTTCTGCTCA TCTTCAGCCT GCTTGTGCAT TGTCCCTTAA TTCATGTTGG GACCATTAGC	548
30	TTCTATGCTG CAAGGCCCGG GTCTGAGCCT AACCGGACTT ATGTTTGTGA CTATGGAAGC	608
	GAGTCAGATT ACAACCCAC CACGGTTCTG TGGTTGGCTC GAGAGACCGA TGGCTCCTGG	668
	ATCTCTGTTT TTTCCGTCA CAACGGCTCC TCAACTGCAG CCCCCGGGGT CGTCGCGCAC	728
	TTTACTGACC ACAACAGCAG CATTGTGGTG CCCAGTATT ACCTCCTCAA CAACTCACTC	788
	TCTAAGCTCT GCTGCTCATA CCGGCACAAC GAGCGTTCTC AGTTTACCTG CAAACAAGCT	848
35	GACGTCCTTA CCTGTCACGA GCCCGGCAAG CCGCTCAGCC TCCGCGTCTC CCCCCGCGTG	908
	GGAAGTCCCC ACCAAGCAGT CACTTGGTTT TTTCAAAATG TACCCATAGC TACTGTTTAC	968
	CGACCTTGGG GCAATGTAAC TTGGTTTTGT CCTCCCTTCA TGTGTACCTT TAATGTCAGC	1028
	CTGAAGTCCC TACTTATTTA CAACTTTTCT GACAAAACCG GGGGGCAATA CACAGCTCTC	1088
	ATGCACTCCG GACCTGCTTC CCTCTTTTCTG CTCTTTAAGC CAACGACTTG TGTACCAAG	1148

	GTGGAGGACC CGCCGTATGC CAACGACCCG GCCTCGCCTG TGTGGCGCCC ACTGCTTTT	1208
	GCCTTCGTCC TCTGCACCG CTGCGCGGTG TTGTTAACCG CCTTCGGTCC ATCGATTCTA	1268
	TCCGGTACCC GAAAGCTTAT CTCAGCCCGC TTTTGGAGTC CCGAGCCCTA TACCACCCTC	1328
	CACATAACAGT CCCCCATGG AGCCAGACGG AGTTCATGCC GAGCAGCAGT TTATCCTCAA	1388
5	TCAGATTTC TGCGCCAACA CTGCCCTCCA GCGTCAAAGG GAGGAAC TAG CTTCCTTGT	1448
	CATGTTGCAT GCCTGTAAGC GTGGCCTCTT TTGTCCAGTC AAAACTTACA AGCTCAGCCT	1508
	CAACGCCCTCG GCCAGCGAGC ACAGCCTGCA CTTTGAAAA AGTCCCTCCC GATTCAACCCT	1568
	GGTCAACACT CACGCCGAG CTCTGTGCG AGTGGCCCTA CACCACCAGG GAGCTTCCGG	1628
	CAGCATCCG TGTCTCTGT CCCACGCCGA GTGCCCTCCC GTCTCTCTCA AGACCTCTG	1688
10	TGCCTTTAAC TTTTATAGATT AGCTGAAAGC AAATATAAAA TGGTGTGCTT ACCGTAATTC	1748
	TGTTTTGACT TGTGTGCTT ATTCTCTCCC CTGCGCCGTA ATCCAGTGCC CCTCTTCAA	1808
	ACTCTCGTAT CCTATGCGAT TCGCATAGGC ATATTTTCTA AAAGCTCTGA AGTCAACATC	1868
	ACTCTCAAAC ACTTCTCCGT TGTAGGTTAC TTTCATCTAC AGATAAAGTC ATCCACCGGT	1928
	TAACATCATG AAGAGAAGTG TGCCCCAGGA CTTAATCTT GTGTATCCGT ACAAGGCTAA	1988
15	GAGGCCCAAC ATCATGCCGC CCTTTTTGA CCGCAATGGC TTTGTTGAAA ACCAAGAAGC	2048
	CACGCTAGCC ATGCTTGTGG AAAAGCCGCT CACGTTCCGAC AAGGAAGGTG CGCTGACCCT	2108
	GGGCGTCGGA CGCGGCATCC GCATTAACCC CGCGGGGCTT CTGGAGACAA ACGACCTCGC	2168
	GTCCGCTGTC TTCCACCGC TGCCCTCCGA TGAGGCCGCG AACGTCACGC TCAACATGTC	2228
	TGACGGGCTA TATACTAAGG ACAACAAGCT AGCTGTCAAA GTAGGTCCCG GGCTGTCCCT	2288
20	CGACTCCAAT AATGCTCTCC AGGTCCACAC AGGCGACGGG CTCACGGTAA CCGATGACAA	2348
	GGTGTCTCTA AATACCCAAG CTCCCTCTC GACCACCAGC GCGGGCCTCT CCCTACTTCT	2408
	GGGTCCAGC CTCCACTTAG GTGAGGAGGA ACGACTAACA GTAAACACCG GAGCGGGCCT	2468
	CCAAATTAGC AATAACGCTC TGCCCGTAAA AGTAGGTTCA GGTATCACCG TAGATGCTCA	2528
	AAACAGCTC GCTGCATCCC TGGGGGACGG TCTAGAAAGC AGAGATAATA AAAGTGTCTG	2588
25	TAAGGCTGGG CCCGGACTTA CAATAACTAA TCAAGCTCTT ACTGTTGCTA CCGGGAACGG	2648
	CCTTCAGGTC AACCCGAAG GGCAACTGCA GCTAACATT ACTGCCGGTC AGGGCCTCAA	2708
	CTTTGCAAAC AACAGCCTCG CCGTGGAGCT GGGCTCGGGC CTGCATTTTC CCCCTGGCCA	2768
	AAACCAAGTA AGCCTTTATC CCGGAGATGG AATAGACATC CGAGATAATA GGTGACTGT	2828
	GCCGCTGGG CCAGGCCTGA GAATGCTCAA CCACCAACTT GCCGTAGCTT CCGGAGACGG	2888
30	TTTAGAAGTC CACAGCGACA CCCTCCGGT AAAGCTCTCC CACGGCCTGA CATTTGAAAA	2948
	TGGCGCCGTA CGAGCAAAAC TAGGACCAGG ACTTGGCACA GACGACTCTG GTCGGTCCGT	3008
	GGTTCGCACA GGTGAGGAC TTAGAGTTGC AAACGGCCAA GTCCAGATCT TCAGCGGAAG	3068
	AGGCACCGCC ATCGGCACTG ATAGCAGCCT CACTCTCAAC ATCCGGGCGC CCCTACAATT	3128
	TTCTGGACCC GCCTTGACTG CTAGTTTGCA AGGCAGTGGT CCGATTACTT ACAACAGCAA	3188
35	CAATGGCACT TTCGGTCTCT CTATAGGCC CGGAATGTGG GTAGACCAA ACAGACTTCA	3248
	GGTAAACCCA GCGCTGGTT TAGTCTTCCA AGGAAACAAC CTTGTCCCAA ACCTTGCGGA	3308
	TCCGCTGGCT ATTTCCGACA GCAAAATTAG TCTCAGTCTC GGTCCCGGCC TGACCAAGC	3368
	TTCCAACGCC CTGACTTTAA GTTTAGGAAA CGGGCTTGAA TTCTCCAATC AAGCGTTGC	3428
	TATAAAGCG GCGCGGGCT TACGCTTTGA GTCTTCTCA CAAGCTTAG AGAGCAGCCT	3488

-75-

CACAGTCGGA AATGGCTTAA CGCTTACCGA TACTGTGATC CGCCCCAACC TAGGGGACGG 3548
 CCTAGAGGTC AGAGACAATA AAATCATTGT TAAGCTGGGC GCGAATCTTC GTTTTGAAAA 3608
 CGGAGCCGTA ACCGCCGGCA CCGTTAACCC TTCTGCGCCC GAGGCACCAC CAACTCTCAC 3668
 TGCAGAACCA CCCCTCCGAG CCTCCAATC CCATCTTCAA CTGTCCCTAT CGGAGGGCTT 3728
 5 GGTGTGTCAT AACACGCCC TTGCTCTCCA ACTGGGAGAC GGCATGGAAG TAAATCAGCA 3788
 CGGACTTACT TTAAGAGTAG GCTCGGGTTT GCAAATGCGT GACGGCATT TAAACAGTTAC 3848
 ACCCAGCGGC ACTCCTATTG AGCCAGACT GACTGCCCA CTGACTCAGA CAGAGAATGG 3908
 AATCGGGCTC GCTCTCGGC CCGGCTTGA ATTAGACGAG AGCGCGCTCC AAGTAAAAGT 3968
 TGGGCCCGGC ATGCGCCTGA ACCCTGTAGA AAAGTATGTA ACCCTGCTCC TGGGTCCTGG 4028
 10 CCTTAGTTTT GGGCAGCCGG CCAACAGGAC AAATTATGAT GTGCGCGTTT CTGTGAGGCC 4088
 CCCCATGTTT TTCGGACAGC GTGGTCAGCT CACATTTTGA GTGGTCACG GACTACACAT 4148
 TCAAAATTC AAATTCAGC TCAATTTGG ACAAGGCCTC AGAACTGACC CCGTCACCAA 4208
 CCAGCTGGA GTGCCCTCG GTCAAGGTTT GGAAATTGCA GACGAATCCC AGGTTAGGGT 4268
 TAAATTGGGC GATGGCCTGC AGTTTGATTC ACAAGCTCG ATCACTACCG CTCCTAACAT 4328
 15 GGTCACTGAA ACTCTGTGA CCGGAACAGG CAGTAATGCT AATGTTACAT GCGGGGGCTA 4388
 CACTGCCCC GGCAGCAAAC TCTTTTGG TCTCACTCG TTCAGCACTG GTCTAGTTTT 4448
 AGGAAACATG ACTATTGACA GCAATGCATC CTTTGGGCAA TACATTAACG CGGGACACGA 4508
 ACAGATCGAA TGCTTTATAT TGTGGACAA TCAGGGTAAC CTAAAAGAAG GATCTAACTT 4568
 GCAAGGCACT TGGGAAGTGA AGAACAAACC CTCTGCTTCC AAAGCTGCTT TTTTGCCTTC 4628
 20 CACCGCCCTA TACCCATCC TCAACGAAAG CCGAGGGAGT CTTCTGGAA AAAATCTTGT 4688
 GGGCATGCAA GCCATACTGG GAGGCGGGGG CACTTGCACT GTGATAGCCA CCCTCAATGG 4748
 CAGACGCAGC AACAACTATC CCGCGGGCCA GTCCATAATT TTCGTGTGGC AAGAATTCAA 4808
 CACCATAGCC CGCCAACCTC TGAACCACTC TACACTTACT TTTTCTTACT GGACTTAAAT 4868
 AAGTTGGAAA TAAAGAGTTA AACTGAATGT TTAAGTGCAA CAGACTTTTA TTGGTTTGG 4928
 25 CTCACAACAA ATTACAACAG CATAGACAAG TCATACCGGT CAAACAACAC AGGCTCTCGA 4988
 AAACGGGCTA ACCGCTCAA GAATCTGTCA CGCAGACGAG CAAGTCCTAA ATGTTTTTTC 5048
 ACTCTCTCG GGGCCAAGTT CAGCATGTAT CGGATTTTCT GCTTACACCT TT 5100

(2) INFORMATION FOR SEQ ID NO:16:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 139 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:
 Leu Ile Lys Gln Pro Val Val Gly Thr Thr His Val Glu Met Pro Arg
 1 5 10 15
 35 Asn Glu Val Leu Glu Gln His Leu Thr Ser His Gly Ala Gln Ile Ala
 20 25 30
 Gly Gly Gly Ala Ala Gly Asp Tyr Phe Lys Ser Pro Thr Ser Ala Arg
 35 40 45
 Thr Leu Ile Pro Leu Thr Ala Ser Cys Leu Arg Pro Asp Gly Val Phe
 50 55 60

-76-

Gln Leu Gly Gly Gly Ser Arg Ser Ser Phe Asn Pro Leu Gln Thr Asp
 65 70 75 80
 Phe Ala Phe His Ala Leu Pro Ser Arg Pro Arg His Gly Gly Ile Gly
 85 90 95
 Ser Arg Gln Phe Val Glu Glu Phe Val Pro Ala Val Tyr Leu Asn Pro
 100 105 110
 5 Tyr Ser Gly Pro Pro Asp Ser Tyr Pro Asp Gln Phe Ile Arg His Tyr
 115 120 125
 Asn Val Tyr Ser Asn Ser Val Ser Gly Tyr Ser
 130 135

(2) INFORMATION FOR SEQ ID NO:17:

10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5100 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

15 (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 408..1331

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CCTCATCAAA CAACCCGTGG TGGGCACCAC CCACGTGGAA ATGCCTCGCA ACGAAGTCCT 60
 AGAACAACAT CTGACCTCAC ATGGCGCTCA AATCGCGGGC GGAGGCGCTG CGGGCGATTA 120
 CTTTAAAGC CCCACTTCAG CTCGAACCTT TATCCGCTC ACCGCCTCCT GCTTAAGACC 180
 20 AGATGGAGTC TTTCAACTAG GAGGAGGCTC GCGTTCATCT TTCAACCCCC TGCAAACAGA 240
 TTTTGCCCTC CACGCCCTGC CCTCCAGACC GCGCCACGGG GGCATAGGAT CCAGGCAGTT 300
 TGTAGAGGAA TTTGTGCCCG CCGTCTACCT CAACCCCTAC TCGGGACCGC CGGACTCTTA 360
 TCCGGACCAG TTTATACGCC ACTACAACGT GTACAGCAAC TCTGTGA GCG GTT ATA 416
 Ala Val Ile
 1
 25 GCT GAG ATT GTA AGA CTC TCC TAT CTG TCT CTG TGC TGC TTT TCC GCT 464
 Ala Glu Ile Val Arg Leu Ser Tyr Leu Ser Leu Cys Cys Phe Ser Ala
 5 10 15
 TCA AGC CCC ACA AGC ATG AAG GGG TTT CTG CTC ATC TTC AGC CTG CTT 512
 Ser Ser Pro Thr Ser Met Lys Gly Phe Leu Leu Ile Phe Ser Leu Leu
 20 25 30 35
 30 GTG CAT TGT CCC CTA ATT CAT GTT GGG ACC ATT AGC TTC TAT GCT GCA 560
 Val His Cys Pro Leu Ile His Val Gly Thr Ile Ser Phe Tyr Ala Ala
 40 45 50
 AGG CCC GGG TCT GAG CCT AAC GCG ACT TAT GTT TGT GAC TAT GGA AGC 608
 Arg Pro Gly Ser Glu Pro Asn Ala Thr Tyr Val Cys Asp Tyr Gly Ser
 55 60 65
 GAG TCA GAT TAC AAC CCC ACC ACG GTT CTG TGG TTG GCT CGA GAG ACC 656
 Glu Ser Asp Tyr Asn Pro Thr Thr Val Leu Trp Leu Ala Arg Glu Thr
 70 75 80
 35 GAT GGC TCC TGG ATC TCT GTT CTT TTC CGT CAC AAC GGC TCC TCA ACT 704
 Asp Gly Ser Trp Ile Ser Val Leu Phe Arg His Asn Gly Ser Ser Thr
 85 90 95
 GCA GCC CCC GGG GTC GTC GCG CAC TTT ACT GAC CAC AAC AGC AGC ATT 752
 Ala Ala Pro Gly Val Val Ala His Phe Thr Asp His Asn Ser Ser Ile
 100 105 110 115

-77-

	GTG GTG CCC CAG TAT TAC CTC CTC AAC AAC TCA CTC TCT AAG CTC TGC Val Val Pro Gln Tyr Tyr Leu Leu Asn Asn Ser Leu Ser Lys Leu Cys 120 125 130	800
	TGC TCA TAC CGG CAC AAC GAG CGT TCT CAG TTT ACC TGC AAA CAA GCT Cys Ser Tyr Arg His Asn Glu Arg Ser Gln Phe Thr Cys Lys Gln Ala 135 140 145	848
5	GAC GTC CCT ACC TGT CAC GAG CCC GGC AAG CCG CTC ACC CTC CGC GTC Asp Val Pro Thr Cys His Glu Pro Gly Lys Pro Leu Thr Leu Arg Val 150 155 160	896
	TCC CCC GCG CTG GGA ACT GCC CAC CAA GCA GTC ACT TGG TTT TTT CAA Ser Pro Ala Leu Gly Thr Ala His Gln Ala Val Thr Trp Phe Phe Gln 165 170 175	944
10	AAT GTA CCC ATA GCT ACT GTT TAC CGA CCT TGG GGC AAT GTA ACT TGG Asn Val Pro Ile Ala Thr Val Tyr Arg Pro Trp Gly Asn Val Thr Trp 180 185 190 195	992
	TTT TGT CCT CCC TTC ATG TGT ACC TTT AAT GTC AGC CTG AAC TCC CTA Phe Cys Pro Phe Met Cys Thr Phe Asn Val Ser Leu Asn Ser Leu 200 205 210	1040
	CTT ATT TAC AAC TTT TCT GAC AAA ACC GGG GGG CAA TAC ACA GCT CTC Leu Ile Tyr Asn Phe Ser Asp Lys Thr Gly Gly Gln Tyr Thr Ala Leu 215 220 225	1088
15	ATG CAC TCC GGA CCT GCT TCC CTC TTT CAG CTC TTT AAG CCA ACG ACT Met His Ser Gly Pro Ala Ser Leu Phe Gln Leu Phe Lys Pro Thr Thr 230 235 240	1136
	TGT GTC ACC AAG GTG GAG GAC CCG CCG TAT GCC AAC GAC CCG GCC TCG Cys Val Thr Lys Val Glu Asp Pro Pro Tyr Ala Asn Asp Pro Ala Ser 245 250 255	1184
20	CCT GTG TGG CGC CCA CTG CTT TTT GCC TTC GTC CTC TGC ACC GGC TGC Pro Val Trp Arg Pro Leu Leu Phe Ala Phe Val Leu Cys Thr Gly Cys 260 265 270 275	1232
	GCG GTG TTG TTA ACC GCC TTC GGT CCA TCG ATT CTA TCC GGT ACC CGA Ala Val Leu Leu Thr Ala Phe Gly Pro Ser Ile Leu Ser Gly Thr Arg 280 285 290	1280
	AAG CTT ATC TCA GCC CGC TTT TGG AGT CCC GAG CCC TAT ACC ACC CTC Lys Leu Ile Ser Ala Arg Phe Trp Ser Pro Glu Pro Tyr Thr Thr Leu 295 300 305	1328
25	CAC TAACAGTCCC CCCATGGAGC CAGACGGAGT TCATGCCGAG CAGCAGTTTA His	1381
	TCCTCAATCA GATTTCCTGC GCCAACACTG CCCTCCAGCG TCAAAGGGAG GAACTAGCTT	1441
	CCCTTGTCAT GTTGTCATGCC TGTAAGCGTG GCCTCTTTTG TCCAGTCAAA ACTTACAAGC	1501
	TCAGCCTCAA CGCCTCGGCC AGCGAGCACA GCCTGCACTT TGAAAAAAGT CCCTCCCGAT	1561
30	TCACCCGGT CAACACTCAC GCCGGAGCTT CTGTGCGAGT GGCCCTACAC CACCAGGGAG	1621
	CTTCCGGCAG CATCCGCTGT TCCTGTTCCC ACGCCGAGTG CCTCCCCGTC CTCCTCAAGA	1681
	CCCTCTGTGC CTTTAACCTT TTAGATTAGC TGAAAGCAAA TATAAATGG TGTGCTTACC	1741
	GTAATTCTGT TTTGACTTGT GTGCTTGATT TCTCCCCCTG CGCCGTAATC CAGTGCCCTT	1801
	CTTCAAACT CTCGTACCCT ATGCCATTGG CATAGGCATA TTTTCTAAAA GCTCTGAAGT	1861
35	CAACATCACT CTCAAACACT TCTCGTTGT AGGTTACTTT CATCTACAGA TAAAGTCATC	1921
	CACCGGTTAA CATCATGAAG AGAAGTGTGC CCCAGGACTT TAATCTTGTG TATCCGTACA	1981
	AGGCTAAGAG GCCCAACATC ATGCCGCCCT TTTTGTACCG CAATGGCTTT GTTGAAAACC	2041
	AAGAAGCCAC GCTAGCCATG CTTGTGGAAG AGCCGCTCAC GTTCGACAAG GAAGGTGCGC	2101
	TGACCCTGGG CGTCGGACGC GGCATCCGCA TTAACCCCGC GGGGCTTCTG GAGACAAACG	2161

-78-

	ACCTCGCGTC CGCTGTCTTC CCACCGCTGG CCTCCGATGA GGCCGGCAAC GTCACGCTCA	2221
	ACATGTCTGA CGGGCTATAT ACTAAGGACA ACAAGCTAGC TGTCAAAGTA GGTCCCGGGC	2281
	TGTCCCTCGA CTCCAATAAT GCTCTCCAGG TCCACACAGG CGACGGGCTC ACGGTAACCG	2341
	ATGACAAAGT GTCTCTAAAT ACCCAAGCTC CCTCTCGAC CACCAGCGCG GGCCTCTCCC	2401
5	TACTTCTGGG TCCCAGCCTC CACTTAGGTG AGGAGGAACG ACTAACAGTA AACACCGGAG	2461
	CGGGCCTCCA AATTAGCAAT AACGCTCTGG CCGTAAAGT AGGTTCAAGT ATCACCCTAG	2521
	ATGCTCAAAA CCAGCTCGCT GCATCCCTGG GGGACGGTCT AGAAAGCAGA GATAATAAAA	2581
	CTGTCTGTTA GGCTGGGCCG GGACTTACAA TAACTAATCA AGCTCTTACT GTTGCTACCG	2641
	GGAACGGCCT TCAGGTCAAC CCGGAAGGGC AACTGCAGCT AAACATTACT GCCGGTCAGG	2701
10	GCCTCAACTT TGCAAACAAC AGCCTCGCCG TGGAGCTGGG CTCGGGCCCTG CATTTTCCCC	2761
	CTGGCCAAAA CCAAGTAAGC CTTTATCCCG GAGATGGAAT AGACATCCGA GATAATAGGG	2821
	TGACTGTGCC CGCTGGGCCA GGCCTGAGAA TGCTCAACCA CCAACTTGCC GTAGCTTCCG	2881
	GAGACGGTTT AGAAGTCCAC AGCGACACCC TCCGGTTAAA GCTCTCCCA GGCCTGACAT	2941
	TTGAAAATGG CGCCGTACGA GCAAACTAG GACCAGGACT TGGCACAGAC GACTCTGGTC	3001
15	GGTCCGTGGT TCGCACAGGT CGAGGACTTA GAGTTGCAAA CGGCCAAGTC CAGATCTTCA	3061
	GCGGAAGAGG CACCGCCATC GGCCTGATA GCAGCCTCAC TCTCAACATC CGGGCGCCCC	3121
	TACAATTTTC TGGACCCGCC TTGACTGCTA GTTTGCAAGG CAGTGGTCCG ATTACTTACA	3181
	ACAGCAACAA TGGCACTTTC GGTCTCTCTA TAGGCCCCGG AATGTGGGA GACCAAAACA	3241
	GACTTCAGGT AAACCCAGGC GCTGGTTTGG TCTTCCAAGG AAACAACCTT GTCCCAAACC	3301
20	TTGCGGATCC GCTGGCTATT TCCGACAGCA AAATTAGTCT CAGTCTCGGT CCCGGCCTGA	3361
	CCCAAGCTTC CAACGCCCTG ACTTTAAGTT TAGGAAACGG GCTTGAATTC TCCAATCAAG	3421
	CCGTTGCTAT AAAAGCGGGC CGGGGCTTAC GCTTTGAGTC TTCCTCACA GCTTTAGAGA	3481
	GCAGCCTCAC AGTCGGAAAT GGCTTAACGC TTACCGATAC TGTGATCCGC CCCAACCTAG	3541
	GGGACGGCCT AGAGGTCAGA GACAATAAAA TCATTGTAA GCTGGGCGCG AATCTTCGTT	3601
25	TTGAAAACGG AGCCGTAACC GCCGGCACCG TTAACCCCTC TGGCCCCGAG GCACCACCAA	3661
	CTCTCACTGC AGAACCACCC CTCGAGCCT CCAACTCCCA TCTTCAACTG TCCCTATCGG	3721
	AGGGCTTGTT TGTGCATAAC AACGCCCTTG CTCTCCAAC TGGGACGGC ATGGAAGTAA	3781
	ATCAGCACGG ACTTACTTTA AGAGTAGGCT CGGGTTTGCA AATGCGTGAC GGCATTTTAA	3841
	CAGTTACACC CAGCGGCACT CCTATTGAGC CCAGACTGAC TGCCCCACTG ACTCAGACAG	3901
30	AGAATGGAAT CGGGCTCGCT CTCGGCGCCG GCTTGGAAAT AGACGAGAGC GCGCTCCAAG	3961
	TAAAAGTTGG GCCCGGCATG CGCCTGAACC CTGTAGAAA GTATGTAACC CTGCTCCTGG	4021
	GTCCTGGCCT TAGTTTTGGG CAGCCGGCCA ACAGGACAAA TTATGATGTG CCGCTTTCTG	4081
	TGGAGCCCCC CATGGTTTTG GGACAGCGTG GTCAGCTCAC ATTTTATAGT GGTACGGAC	4141
	TACACATTCA AAATTCCAAA CTTCACTCA ATTTGGGACA AGGCCTCAGA ACTGACCCCG	4201
35	TCACCAACCA GCTGGAAGTG CCCCTCGGTC AAGTTTGGG AATTGACAGC GAATCCAGG	4261
	TTAGGGTTAA ATTTGGCGAT GGCCTGCAGT TTGATTCACA AGCTCGCATC ACTACCGCTC	4321
	CTAACATGGT CACTGAACT CTGTGGACCG GAACAGGCAG TAATGCTAAT GTTACATGGC	4381
	GGGGCTACAC TGCCCCCGGC AGCAAACTCT TTTTGAGTCT CACTCGGTC AGCACTGGTC	4441
	TAGTTTTAGG AAACATGACT ATTGACAGCA ATGCATCCTT TGGGCAATAC ATTAACGCGG	4501

-79-

GACACGAACA GATCGAATGC TTTATATTGT TGGACAATCA GGGTAACCTA AAAGAAGGAT 4561
 CTAACCTTGA AGGCACCTGG GAAGTGAAGA ACAACCCCTC TGCTTCCAAA GCTGCTTTTT 4621
 TGCCTTCCAC CGCCCTATAC CCCATCCTCA ACGAAAGCCG AGGGAGTCTT CCTGGAAAAA 4681
 ATCTTGTTGG CATGCAAGCC ATACTGGGAG GCGGGGGCAC TTGCACTGTG ATAGCCACCC 4741
 5 TCAATGGCAG ACGCAGCAAC AACTATCCCG CGGGCCAGTC CATAATTTTC GTGTGGCAAG 4801
 AATTCAACAC CATAGCCCGC CAACCTCTGA ACCACTCTAC ACTTACTTTT TCTTACTGGA 4861
 CTTAAATAAG TTGGAATAA AGAGTTAAAC TGAATGTTA AGTGCAACAG ACTTTTATTG 4921
 GTTTTGGCTC ACAACAAATT ACAACAGCAT AGACAAGTCA TACCGGTCAA ACAACACAGG 4981
 CTCTCGAAAA CGGGCTAACC GCTCCAAGAA TCTGTACGCG AGACGAGCAA GTCCTAAATG 5041
 10 TTTTTCCTACT CTCTTCGGGG CCAAGTTCAG CATGTATCGG ATTTTCTGCT TACACCTTT 5100

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 308 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Ala Val Ile Ala Glu Ile Val Arg Leu Ser Tyr Leu Ser Leu Cys Cys
 1 5 10 15
 Phe Ser Ala Ser Ser Pro Thr Ser Met Lys Gly Phe Leu Leu Ile Phe
 20 25 30
 20 Ser Leu Leu Val His Cys Pro Leu Ile His Val Gly Thr Ile Ser Phe
 35 40 45
 Tyr Ala Ala Arg Pro Gly Ser Glu Pro Asn Ala Thr Tyr Val Cys Asp
 50 55 60
 Tyr Gly Ser Glu Ser Asp Tyr Asn Pro Thr Thr Val Leu Trp Leu Ala
 65 70 75 80
 25 Arg Glu Thr Asp Gly Ser Trp Ile Ser Val Leu Phe Arg His Asn Gly
 85 90 95
 Ser Ser Thr Ala Ala Pro Gly Val Val Ala His Phe Thr Asp His Asn
 100 105 110
 Ser Ser Ile Val Val Pro Gln Tyr Tyr Leu Leu Asn Asn Ser Leu Ser
 115 120 125
 Lys Leu Cys Cys Ser Tyr Arg His Asn Glu Arg Ser Gln Phe Thr Cys
 130 135 140
 30 Lys Gln Ala Asp Val Pro Thr Cys His Glu Pro Gly Lys Pro Leu Thr
 145 150 155 160
 Leu Arg Val Ser Pro Ala Leu Gly Thr Ala His Gln Ala Val Thr Trp
 165 170 175
 Phe Phe Gln Asn Val Pro Ile Ala Thr Val Tyr Arg Pro Trp Gly Asn
 180 185 190
 35 Val Thr Trp Phe Cys Pro Pro Phe Met Cys Thr Phe Asn Val Ser Leu
 195 200 205
 Asn Ser Leu Leu Ile Tyr Asn Phe Ser Asp Lys Thr Gly Gly Gln Tyr
 210 215 220
 Thr Ala Leu Met His Ser Gly Pro Ala Ser Leu Phe Gln Leu Phe Lys
 225 230 235 240

-80-

Pro Thr Thr Cys Val Thr Lys Val Glu Asp Pro Pro Tyr Ala Asn Asp
245 250 255

Pro Ala Ser Pro Val Trp Arg Pro Leu Leu Phe Ala Phe Val Leu Cys
260 265 270

Thr Gly Cys Ala Val Leu Leu Thr Ala Phe Gly Pro Ser Ile Leu Ser
275 280 285

5 Gly Thr Arg Lys Leu Ile Ser Ala Arg Phe Trp Ser Pro Glu Pro Tyr
290 295 300

Thr Thr Leu His
305

(2) INFORMATION FOR SEQ ID NO:19:

10 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5100 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

15 (ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 529..954

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

	CCTCATCAA CAACCCGTGG TGGGCACCAC CCACGTGGAA ATGCCTCGCA ACGAAGTCCT	60
	AGAACAACAT CTGACCTCAC ATGGCGCTCA AATCGCGGGC GGAGGCGCTG CGGGCGATTA	120
	CTTTAAAGC CCCACTTCAG CTCGAACCCCT TATCCCGCTC ACCGCCCTCT GCTTAAGACC	180
20	AGATGGAGTC TTTCAACTAG GAGGAGGCTC GCGTTCATCT TTCAACCCCC TGCAACAGA	240
	TTTTGCCTTC CACGCCCTGC CCTCCAGACC GCGCCACGGG GGCATAGGAT CCAGGCAGTT	300
	TGTAGAGGAA TTTGTGCCG CCGTCTACCT CAACCCCTAC TCGGGACCGC CGGACTCTTA	360
	TCCGGACCAG TTTATACGCC ACTACAACGT GTACAGCAAC TCTGTGAGCG GTTATAGCTG	420
	AGATTGTAAG ACTCTCCTAT CTGTCTCTGT GCTGCTTTTC CGCTTCAAGC CCCACAAGCA	480
25	TGAAGGGGTT TCTGCTCATC TTCAGCCTGC TTGTGCATTG TCCCCTAA TTC ATG TTG	537
	Phe Met Leu 1	
	GGA CCA TTA GCT TCT ATG CTG CAA GGC CCG GGT CTG AGC CTA ACG CGA	585
	Gly Pro Leu Ala Ser Met Leu Gln Gly Pro Gly Leu Ser Leu Thr Arg 5 10 15	
30	CTT ATG TTT GTG ACT ATG GAA GCG AGT CAG ATT ACA ACC CCA CCA CGG	633
	Leu Met Phe Val Thr Met Glu Ala Ser Gln Ile Thr Thr Pro Pro Arg 20 25 30 35	
	TTC TGT GGT TGG CTC GAG AGA CCG ATG GCT CCT GGA TCT CTG TTC TTT	681
	Phe Cys Gly Trp Leu Glu Arg Pro Met Ala Pro Gly Ser Leu Phe Phe 40 45 50	
	TCC GTC ACA ACG GCT CCT CAA CTG CAG CCC CCG GGG TCG TCG CGC ACT	729
	Ser Val Thr Thr Ala Pro Gln Leu Gln Pro Pro Gly Ser Ser Arg Thr 55 60 65	
35	TTA CTG ACC ACA ACA GCA GCA TTG TGG TGC CCC AGT ATT ACC TCC TCA	777
	Leu Leu Thr Thr Thr Ala Ala Leu Trp Cys Pro Ser Ile Thr Ser Ser 70 75 80	
	ACA ACT CAC TCT CTA AGC TCT GCT GCT CAT ACC GGC ACA ACG AGC GTT	825
	Thr Thr His Ser Leu Ser Ser Ala Ala His Thr Gly Thr Thr Ser Val 85 90 95	

-81-

	CTC AGT TTA CCT GCA AAC AAG CTG ACG TCC CTA CCT GTC ACG AGC CCG Leu Ser Leu Pro Ala Asn Lys Leu Thr Ser Leu Pro Val Thr Ser Pro 100 105 110 115	873
	GCA AGC CGC TCA CCC TCC GCG TCT CCC CCG CGC TGG GAA CTG CCC ACC Ala Ser Arg Ser Pro Ser Ala Ser Pro Pro Arg Trp Glu Leu Pro Thr 120 125 130	921
5	AAG CAG TCA CTT GGT TTT TTC AAA ATG TAC CCA TAGCTACTGT TTACCGACCT Lys Gln Ser Leu Gly Phe Phe Lys Met Tyr Pro 135 140	974
	TGGGGCAATG TAACCTGGTT TTGCTCTCCC TTCATGTGTA CCTTTAATGT CAGCCTGAAC	1034
	TCCCTACTTA TTTACAACCT TTCTGACAAA ACCGGGGGGC AATACACAGC TCTCATGCAC	1094
	TCCGGACCTG CTTCCCTCTT TCAGCTCTTT AAGCCAACGA CTGTGTGCAC CAAGGTGGAG	1154
10	GACCCGCCGT ATGCCAACGA CCCGGCCTCG CCTGTGTGGC GCCCACTGCT TTTTGCCTTC	1214
	GTCTCTGCA CCGGCTGCGC GGTGTTGTTA ACCGCCTTCG GTCCATCGAT TCTATCCGCT	1274
	ACCCGAAAGC TTATCTCAGC CCGCTTTTGG AGTCCCGAGC CCTATACCAC CCTCCACTAA	1334
	CAGTCCCCC ATGGAGCCAG ACGGAGTTCA TGCCGAGCAG CAGTTTATCC TCAATCAGAT	1394
	TTCTGCGCC AACACTGCCC TCCAGCGTCA AAGGGAGGAA CTAGCTTCCC TTGTCATGTT	1454
15	GCATGCCTGT AAGCGTGGCC TCTTTGTCC AGTCAAACT TACAAGTCA GCCTCAACGC	1514
	CTCGGCCAGC GAGCACAGCC TGCACCTTGA AAAAAGTCCC TCCCGATTCA CCCTGGTCAA	1574
	CACTCACGCC GGAGCTTCTG TGCGAGTGGC CTTACACCAC CAGGGAGCTT CCGGCAGCAT	1634
	CCGCTGTTCC TGTTCCACG CCGAGTGCT CCCCGTCTC CTCAAGACC TCTGTGCTT	1694
	TAACTTTTTA GATTAGTGA AAGCAAATAT AAAATGGTGT GCTTACCGTA ATTCTGTTTT	1754
20	GACTTGTGTG CTTGATTCT CCCCTGCGC CGTAATCCAG TGCCCTCTT CAAACTCTC	1814
	GTACCTATG CGATTGCGAT AGGCATATTT TCTAAAAGCT CTGAAGTCAA CATCACTCTC	1874
	AAACACTTCT CCGTTGTAGG TTACTTTCAT CTACAGATAA AGTCATCCAC CGGTTAACAT	1934
	CATGAAGAGA AGTGTGCCCC AGGACTTTAA TCTGTGTAT CCGTACAAGG CTAAGAGGCC	1994
	CAACATCATG CCGCCCTTTT TTGACCGCAA TGGCTTTGTT GAAAACCAAG AAGCCACGCT	2054
25	AGCCATGCTT GTGGAAAAGC CGCTCACGTT CGACAAGGAA GGTGCGCTGA CCCTGGGCGT	2114
	CGGACGGGGC ATCEGCATTA ACCCGCGGG GCTTCTGGAG ACAAACGACC TCGGCTCCGC	2174
	TGTCTTCCCA CCGCTGGCCT CCGATGAGGC CGGCAACGTC ACGCTCAACA TGTCTGACGG	2234
	GCTATATACT AAGGACAACA AGCTAGCTGT CAAAGTAGGT CCCGGGCTGT CCCTCGACTC	2294
	CAATAATGCT CTCCAGGTCC ACACAGGCGA CGGGCTCACG GTAACCGATG ACAAGGTGTC	2354
30	TCTAAATACC CAAGCTCCCC TCTCGACCAC CAGCGCGGGC CTCTCCCTAC TTCTGGGTCC	2414
	CAGCCTCCAC TTAGGTGAGG AGGAACGACT AACAGTAAAC ACCGGAGCGG GCCTCCAAAT	2474
	TAGCAATAAC GCTCTGGCCG TAAAAGTAGG TTCAGGTATC ACCGTAGATG CTCAAAACCA	2534
	GCTCGCTGCA TCCCTGGGGG ACGGTCTAGA AAGCAGAGAT AATAAACTG TCGTTAAGGC	2594
	TGGGCCCCGA CTTACAATAA CTAATCAAGC TCTTACTGTT GCTACCGGGA ACGGCCTTCA	2654
35	GGTCAACCCG GAAGGGCAAC TGCAGCTAAA CATTACTGCC GGTGAGGGCC TCAACTTTGC	2714
	AAACAACAGC CTCGCCGTGG AGCTGGGCTC GGGCCTGCAT TTTCCCCCTG GCCAAAACCA	2774
	AGTAAGCCTT TATCCCGGAG ATGGAATAGA CATCCGAGAT AATAGGTGA CTGTGCCCCG	2834
	TGGGCCAGGC CTGAGAATGC TCAACCACCA ACTTGCCGTA GCTTCCGGAG ACGGTTTAGA	2894
	AGTCCACAGC GACACCCTCC GGTAAAGCT CTCCCACGGC CTGACATTTG AAAATGGCGC	2954

	CGTACGAGCA AACTAGGAC CAGGACTGG CACAGACGAC TCTGGTCGGT CCGTGGTTCG	3014
	CACAGGTCGA GGACTIONAG ITGCAACGG CCAAGTCCAG ATCTTCAGCG GAAGAGGCAC	3074
	CGCCATCGGC ACTGATAGCA GCCTCACTCT CAACATCCGG GCGCCCTAC AATTTCTGG	3134
	ACCCGCCCTG ACTGCTAGTT TGCAAGGCAG TGGTCCGATT ACTTACAACA GCAACAATGG	3194
5	CACITTCGGT CTCTCTATAG GCCCCGGAAT GTGGGTAGAC CAAAACAGAC TTCAGGTAAA	3254
	CCCAGGCGCT GGTITAGTCT TCCAAGGAAA CAACCTTGTC CCAACCTTG CGGATCCGCT	3314
	GGCTATTTCC GACAGCAAAA TTAGTCTCAG TCTCGGTCCC GGCCTGACCC AAGCTTCCAA	3374
	CGCCCTGACT TTAAGTTTAG GAAACGGGCT TGAATTCTCC AATCAAGCCG TTGTATAAA	3434
	AGCGGGCCGG GGCTTACGCT TTGAGTCTTC CTCACAAGCT TTAGAGAGCA GCCTCACAGT	3494
10	CGGAATGGC TTAACGCTTA CCGATACTGT GATCCGCCCC AACCTAGGGG ACGGCCTAGA	3554
	GGTCAGAGAC AATAAAATCA TTGTTAAGCT GGGCGCGAAT CTTCGTTTTG AAAACGGAGC	3614
	CGTAACCGCC GGCACCGTTA ACCCTTCTGC GCGCGAGGCA CCACCAACTC TCACTGCAGA	3674
	ACCACCCCTC CGAGCCTCCA ACTCCCATCT TCAACTGTCC CTATCGGAGG GCTTGGTTGT	3734
	GCATAACAAC GCCCTTGCTC TCCAACCTGG AGACGGCATG GAAGTAAATC AGCACGGACT	3794
15	TACTTTAAGA GTAGGCTCGG GTTTGCAAAT GCGTGACGGC ATTTTAACAG TTACACCCAG	3854
	CGGCACCTCT ATTGAGCCCA GACTGACTGC CCCACTGACT CAGACAGAGA ATGGAATCGG	3914
	GCTCGCTCTC GCGCGCGGCT TGGAAATAGA CGAGAGCGCG CTCCAAGTAA AAGTTGGGCC	3974
	CGGCATGCGC CTGAACCCTG TAGAAAAGTA TGTAACCCTG CTCCTGGGTC CTGGCCTTAG	4034
	TTTTGGGCAG CCGGCCAACA GGACAAATTA TGATGTGCGC GTTCTGTGG AGCCCCCAT	4094
20	GGTTTTCGGA CAGCGTGGTC AGCTCACATT TTTAGTGGGT CACGGACTAC ACATTCAAAA	4154
	TTCCAAACTT CAGCTCAATT TGGGACAAGG CCTCAGAACT GACCCCGTCA CCAACCAGCT	4214
	GGAAGTGCCC CTCGGTCAAG GTTTGGAAAT TGCAGACGAA TCCCAGGTTA GGGTTAAAT	4274
	GGGCGATGGC CTGCAGTTTG ATTCACAAGC TCGCATCACT ACCGCTCCTA ACATGGTCAC	4334
	TGAAACTCTG TGGACCGGAA CAGGCAGTAA TGCTAATGTT ACATGGCGGG GCTACACTGC	4394
25	CCCCGGCAGC AAACCTTTTT TGAGTCTCAC TCGGTTGAGC ACTGGTCTAG TTTTAGGAAA	4454
	CATGACTATT GACAGCAATG CATCCTTTGG GCAATACATT AACGCGGGAC ACGAACAGAT	4514
	CGAATGCTTT ATATTGTTGG ACAATCAGGG TAACCTAAAA GAAGGATCTA ACTTGCAAGG	4574
	CACITGGGAA GTGAAGAACA ACCCTCTGC TTCCAAAGCT GCTTTTTTGC CTTCACCGC	4634
	CCTATACCCC ATCCTCAACG AAAGCCGAGG GAGTCTTCT GGAATAATC TTGTGGCAT	4694
30	GCAAGCCATA CTGGGAGGCG GGGGCACTTG CACTGTGATA GCCACCTCA ATGGCAGACG	4754
	CAGCAACAAC TATCCCGCGG GCCAGTCCAT AATTTTCGTG TGGCAAGAAT TCAACACCAT	4814
	AGCCCGCCAA CCTCTGAACC ACTCTACACT TACTTTTTCT TACTGGACTT AAATAAGTTG	4874
	GAAATAAAGA GTTAACTGA ATGTTTAAGT GCAACAGACT TTTATTGGTT TTGGCTCACA	4934
	ACAAATTACA ACAGCATAGA CAAGTCATAC CGGTCAAACA ACACAGGCTC TCGAAAACGG	4994
35	GCTAACCGCT CCAAGAATCT GTCACGAGA CGAGCAAGTC CTAAATGTTT TTTCACTCTC	5054
	TTCCGGGCCA AGTTCAGCAT GTATCGGATT TTCTGCTTAC ACCTTT	5100

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 142 amino acids

-83-

(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

5 Phe Met Leu Gly Pro Leu Ala Ser Met Leu Gln Gly Pro Gly Leu Ser
1 5 10 15
Leu Thr Arg Leu Met Phe Val Thr Met Glu Ala Ser Gln Ile Thr Thr
20 25 30
Pro Pro Arg Phe Cys Gly Trp Leu Glu Arg Pro Met Ala Pro Gly Ser
35 40 45
Leu Phe Phe Ser Val Thr Thr Ala Pro Gln Leu Gln Pro Pro Gly Ser
50 55 60
10 Ser Arg Thr Leu Leu Thr Thr Thr Ala Ala Leu Trp Cys Pro Ser Ile
65 70 75 80
Thr Ser Ser Thr Thr His Ser Leu Ser Ser Ala Ala His Thr Gly Thr
85 90 95
Thr Ser Val Leu Ser Leu Pro Ala Asn Lys Leu Thr Ser Leu Pro Val
100 105 110
15 Thr Ser Pro Ala Ser Arg Ser Pro Ser Ala Ser Pro Pro Arg Trp Glu
115 120 125
Leu Pro Thr Lys Gln Ser Leu Gly Phe Phe Lys Met Tyr Pro
130 135 140

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5100 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1246..1707

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

20 CCTCATCAAA CAACCCGTGG TGGGCACCAC CCACGTGGAA ATGCCTCGCA ACGAAGTCCT 60
AGAACAACAT CTGACCTCAC ATGGCGCTCA AATCGCGGGC GGAGGCGCTG CGGGCGATTA 120
CTTTAAAGC CCCACTTCAG CTCGAACCTT TATCCCGCTC ACCGCTCTCT GCTTAAGACC 180
30 AGATGGAGTC TTTCAACTAG GAGGAGGCTC GCGTTCATCT TTCAACCCCC TGCAAACAGA 240
TTTTGCCTTC CAGGCCCTGC CCTCCAGACC GCGCCACGGG GGCATAGGAT CCAGGCAGTT 300
TGTAGAGGAA TTTGTGCCCG CCGTCTACCT CAACCCCTAC TCGGGACCGC CGGACTCTTA 360
TCCGGACCAG TTTATACGCC ACTACAACGT GTACAGCAAC TCTGTAGCGG GTTATAGCTG 420
AGATTGTAAG ACTCTCTAT CTGTCTCTGT GCTGCTTTTC CGCTTCAAGC CCCACAAGCA 480
35 TGAAGGGGTT TCTGCTCATC TTCAGCCTGC TTGTGCATTG TCCCTAATT CATGTTGGGA 540
CCATTAGCTT CTATGCTGCA AGGCCCCGGT CTGAGCCTAA CGCGACTTAT GTTGTGACT 600
ATGGAAGCGA GTCAGATTAC AACCCACCA CGGTTCTGTG GTTGGCTCGA GAGACCGATG 660
GCTCCTGGAT CTCTGTTCTT TTCCGTCACA ACGGCTCCTC AACTGCAGCC CCCGGGGTGC 720
TCGCGCACTT TACTGACCAC AACAGCAGCA TTGTGGTGCC CCAGTATTAC CTCCTCAACA 780

	ACTCACTCTC TAAGCTCTGC TGCTCATACC GGCACAACGA GCGTTCTCAG TTTACCTGCA	840
	AACAAGCTGA CGTCCCTACC TGTCACGAGC CCGGCAAGCC GCTCACCCCTC CGCGTCTCCC	900
	CCGCGCTGGG AACTGCCCAC CAAGCAGTCA CTTGGTTTTT TCAAAATGTA CCCATAGTA	960
	CTGTTTACCG ACCTTGGGGC AATGTAACCT GGTTTTGTC TCCCTTCATG TGTACCTTTA	1020
5	ATGTCAGCCT GAACTCCCTA CTTATTTACA ACTTTTCTGA CAAAACCGGG GGGCAATACA	1080
	CAGCTCTCAT GCACTCCGGA CCGTCTTCCC TCTTCAGCT CTTTAAGCCA ACGACTTGTG	1140
	TCACCAAGGT GGAGGACCCG CCGTATGCCA ACGACCCGGC CTCGCTGTG TGGCGCCAC	1200
	TGCTTTTTCG CTTCGTCTC TGCACCGCT GCGCGGTGT GTTAA CCG CCT TCG	1254
	Pro Pro Ser	
	1	
10	GTC CAT CGA TTC TAT CCG GTA CCC GAA AGC TTA TCT CAG CCC GCT TTT	1302
	Val His Arg Phe Tyr Pro Val Pro Glu Ser Leu Ser Gln Pro Ala Phe	
	5 10 15	
	GGA GTC CCG AGC CCT ATA CCA CCC TCC ACT AAC AGT CCC CCC ATG GAG	1350
	Gly Val Pro Ser Pro Ile Pro Pro Ser Thr Asn Ser Pro Pro Met Glu	
	20 25 30 35	
	CCA GAC GGA GTT CAT GCC GAG CAG CAG TTT ATC CTC AAT CAG ATT TCC	1398
	Pro Asp Gly Val His Ala Glu Gln Gln Phe Ile Leu Asn Gln Ile Ser	
15	40 45 50	
	TGC GCC AAC ACT GCC CTC CAG CGT CAA AGG GAG GAA CTA GCT TCC CTT	1446
	Cys Ala Asn Thr Ala Leu Gln Arg Gln Arg Glu Glu Leu Ala Ser Leu	
	55 60 65	
	GTC ATG TTG CAT GCC TGT AAG CGT GGC CTC TTT TGT CCA GTC AAA ACT	1494
	Val Met Leu His Ala Cys Lys Arg Gly Leu Phe Cys Pro Val Lys Thr	
	70 75 80	
20	TAC AAG CTC AGC CTC AAC GCC TCG GCC AGC GAG CAC AGC CTG CAC TTT	1542
	Tyr Lys Leu Ser Leu Asn Ala Ser Ala Ser Glu His Ser Leu His Phe	
	85 90 95	
	GAA AAA AGT CCC TCC CGA TTC ACC CTG GTC AAC ACT CAC GCC GGA GCT	1590
	Glu Lys Ser Pro Ser Arg Phe Thr Leu Val Asn Thr His Ala Gly Ala	
	100 105 110 115	
	TCT GTG CGA GTG GCC CTA CAC CAC CAG GGA GCT TCC GGC AGC ATC CGC	1638
	Ser Val Arg Val Ala Leu His His Gln Gly Ala Ser Gly Ser Ile Arg	
25	120 125 130	
	TGT TCC TGT TCC CAC GCC GAG TGC CTC CCC GTC CTC CTC AAG ACC CTC	1686
	Cys Ser Cys Ser His Ala Glu Cys Leu Pro Val Leu Leu Lys Thr Leu	
	135 140 145	
	TGT GCC TTT AAC TTT TTA GAT TAGCTGAAAG CAAATATAAA ATGGTGTGCT	1737
	Cys Ala Phe Asn Phe Leu Asp	
	150	
30	TACCGTAATT CTGTTTGGAC TTGTGTGCTT GATTTCTCCC CCGCGCCGT AATCCAGTGC	1797
	CCCTCTTCAA AACTCTCGTA CCCTATGCGA TTCGCATAGG CATATTTTCT AAAAGCTCTG	1857
	AAGTCAACAT CACTCTCAA CACTTCTCCG TTGTAGGTTA CTTTCATCTA CAGATAAAGT	1917
	CATCCACCGG TTAACATCAT GAAGAGAAGT GTGCCCCAGG ACTTTAATCT TGTGTATCCG	1977
	TACAAGGCTA AGAGGCCCAA CATCATGCCG CCCTTTTTTG ACCGCAATGG CTTTGTGGAA	2037
35	AACCAAGAAG CCACGCTAGC CATGCTGTG GAAAAGCCGC TCACGTTCTGA CAAGGAAGGT	2097
	GCGCTGACCC TGGGCGTCGG ACGCGGCATC CGCATTAAAC CCGCGGGGCT TCTGGAGACA	2157
	AACGACCTCG CGTCCGCTGT CTTCACCG CTGGCTCCG ATGAGGCCGG CAACGTCAGG	2217
	CTCAACATGT CTGACGGGCT ATATACTAAG GACAACAAGC TAGCTGTCAA AGTAGGTCCC	2277
	GGGCTGTCCC TCGACTCCAA TAATGCTCTC CAGGTCCACA CAGGCGACGG GCTCACGGTA	2337

	ACCGATGACA AGGTGTCTCT AAATACCCAA GTCCTCTCT CGACCACCAG CGCGGGCCTC	2397
	TCCCTACTTC TGGGTCCCAG CCTCCACTTA GGTGAGGAGG AACGACTAAC AGTAAACACC	2457
	GGAGCGGGCC TCCAAATTAG CAATAACGCT CTGGCCGTAA AAGTAGGTC AGGTATCACC	2517
	GTAGATGCTC AAAACCAGCT CGCTGCATCC CTGGGGGACG GTCTAGAAAG CAGAGATAAT	2577
5	AAACTGTCTG TTAAGGCTGG GCCCGGACTT ACAATACTA ATCAAGCTCT TACTGTTGCT	2637
	ACCGGGAACG GCCTTCAGGT CAACCCGGAA GGGCAACTGC AGCTAAACAT TACTGCCGGT	2697
	CAGGGCCTCA ACTTTGCAA CAACAGCCTC GCCGTGGAGC TGGGCTCGGG CCTGCATTTT	2757
	CCCCCTGGCC AAAACCAAGT AAGCCTTTAT CCCGGAGATG GAATAGACAT CCGAGATAAT	2817
	AGGGTGACTG TGCCCGCTGG GCCAGGCTG AGAATGCTCA ACCACCACT TGCCGTAGCT	2877
10	TCCGGAGACG GTTTAGAAAT CCACAGCGAC ACCCTCCGGT TAAAGCTCTC CCACGGCCTG	2937
	ACATTTGAAA ATGGCGCCGT ACGAGCAAAA CTAGGACCAG GACTTGGCAC AGACGACTCT	2997
	GGTCGGTCCG TGGTTCGCAC AGGTCGAGGA CTTAGAGTTG CAAACGGCCA AGTCCAGATC	3057
	TTCAGCGGAA GAGGCACCGC CATCGGCACT GATAGCAGCC TCACTCTCAA CATCCGGCG	3117
	CCCTACAAT TTCTGGACC CGCCTTGACT GCTAGTTTGC AAGGCAGTGG TCCGATTACT	3177
15	TACAACAGCA ACAATGGCAC TTTCGGTCTC TCTATAGGCC CCGGAATGTG GGTAGACCAA	3237
	AACAGACTTC AGGTAAACCC AGGCGCTGGT TTAGTCTTCC AAGGAAACAA CCTTGTCCTA	3297
	AACCTTGCGG ATCCGCTGGC TATTTCCGAC AGCAAAATTA GTCTCAGTCT CGGTCCCGGC	3357
	CTGACCCAAG CTTCACACG CCTGACTTTA AGTTTAGGAA ACGGGCTTGA ATTCTCCAAT	3417
	CAGCCGTTG CTATAAAGC GGGCCGGGGC TTACGCTTTG AGTCTTCTC ACAAGCTTTA	3477
20	GAGAGCAGCC TCACAGTCGG AATGGETTA ACGCTTACCG ATACTGTGAT CCGCCCCAAC	3537
	CTAGGGGACG GCCTAGAGGT CAGAGACAAT AAAATCATG TTAAGCTGGG CGCGAATCTT	3597
	CGTTTTGAAA ACGGAGCCGT AACC GCCGGC ACCGTTAACC CTTCTGCGCC CGAGGCACCA	3657
	CCAACCTCA CTGCAGAAC ACCCTCCGA GCCTCCAAC CCCATCTTCA ACTGTCCCTA	3717
	TCCGAGGGCT TGGTTGTGCA TAACAACGCC CTTGCTCTCC AACTGGGAGA CGGCATGGAA	3777
25	GTAAATCAGC ACGGACTTAC TTTAAGAGTA GGCTCGGGT TGCAAATGCG TGACGGCATT	3837
	TTAACAGTTA CACCCAGCGG CACTCCTATT GAGCCAGAC TGACTGCCCC ACTGACTCAG	3897
	ACAGAGAATG GAATCGGGCT CGCTCTCGG GCCGGCTTGG AATTAGACGA GAGCGCGCTC	3957
	CAAGTAAAG TTGGGCCCG CATGCGCCTG AACCCTGTAG AAAAGTATGT AACCTGCTC	4017
	CTGGGTCTG GCCTTAGTTT TGGGCAGCCG GCCAACAGGA CAAATTATGA TGTGCGCGTT	4077
30	TCTGTGAGC CCCCCATGGT TTTCGGACAG CGTGGTCAGC TCACATTTT AGTGGGTAC	4137
	GGACTACACA TTCAAAATTC CAAACTTCAG CTCAATTTG GACAAGGCCT CAGAACTGAC	4197
	CCCGTCACCA ACCAGCTGGA AGTGCCCTC GTCAAGGTT TGGAAATTGC AGACGAATCC	4257
	CAGGTAGGG TTAAATTGGG CGATGGCCTG CAGTTTGATT CACAAGCTCG CATCACTACC	4317
	GCTCCTAACA TGGTCACTGA AACTCTGTG ACCGGAACAG GCAGTAATGC TAATGTTACA	4377
35	TGGCGGGCT AACTGCCCC CGGCAGCAA CTCTTTTGA GTCTCACTCG GTTCAGCACT	4437
	GGTCTAGTT TAGGAAACAT GACTATTGAC AGCAATGCAT CCTTTGGGCA ATACATTAA	4497
	GCGGGACACG AACAGATCGA ATGCTTTATA TTGTTGGACA ATCAGGGTAA CCTAAAGAA	4557
	GGATCTAACT TGCAAGGCAC TTGGGAAGTG AAGAACAACC CCTCTGCTTC CAAAGCTGCT	4617
	TTTTTGCTT CCACCGCCT ATACCCATC CTCAACGAAA GCCGAGGGAG TCTTCTGGA	4677

-86-

AAAAATCTTG TGGGCATGCA AGCCATACTG GGAGGCGGGG GCACTTGAC TGTGATAGCC 4737
 ACCCTCAATG GCAGACGCAG CAACAACATAT CCCGCGGGCC AGTCCATAAT TTTCGTGTGG 4797
 CAAGAATTCA ACACCATAGC CCGCCAACCT CTGAACCACT CTACACTTAC TTTTCTTAC 4857
 TGGACTTAAA TAAGTTGGAA ATAAAGAGTT AACTGAATG TTTAAGTGCA ACAGACTTTT 4917
 5 ATTGGTTTTG GCTCACAACA AATTACAACA GCATAGACAA GTCATACCGG TCAAACAACA 4977
 CAGGCTCTCG AAAACGGGCT AACCGCTCCA AGAATCTGTC ACGCAGACGA GCAAGTCCTA 5037
 AATGTTTTTT CACTCTCTTC GGGGCCAAGT TCAGCATGTA TCGGATTTTC TGCTTACACC 5097
 TTT 5100

(2) INFORMATION FOR SEQ ID NO:22:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 154 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

15

Pro Pro Ser Val His Arg Phe Tyr Pro Val Pro Glu Ser Leu Ser Gln
 1 5 10 15

Pro Ala Phe Gly Val Pro Ser Pro Ile Pro Pro Ser Thr Asn Ser Pro
 20 25 30

Pro Met Glu Pro Asp Gly Val His Ala Glu Gln Gln Phe Ile Leu Asn
 35 40 45

20

Gln Ile Ser Cys Ala Asn Thr Ala Leu Gln Arg Gln Arg Glu Glu Leu
 50 55 60

Ala Ser Leu Val Met Leu His Ala Cys Lys Arg Gly Leu Phe Cys Pro
 65 70 75 80

Val Lys Thr Tyr Lys Leu Ser Leu Asn Ala Ser Ala Ser Glu His Ser
 85 90 95

Leu His Phe Glu Lys Ser Pro Ser Arg Phe Thr Leu Val Asn Thr His
 100 105 110

25

Ala Gly Ala Ser Val Arg Val Ala Leu His His Gln Gly Ala Ser Gly
 115 120 125

Ser Ile Arg Cys Ser Cys Ser His Ala Glu Cys Leu Pro Val Leu Leu
 130 135 140

Lys Thr Leu Cys Ala Phe Asn Phe Leu Asp
 145 150

30

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5100 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

35

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1439..1702

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CCTCATCAAA CAACCCGTGG TGGGACCAC CCACGTGGAA ATGCCTCGCA ACGAAGTCCT 60

-87-

	AGAACAACAT CTGACCTCAC ATGGCGCTCA AATCGCGGGC GGAGGCGCTG CGGGCGATTA	120
	CTTTAAAGC CCCACTTCAG CTCGAACCCCT TATCCCGCTC ACCGCCCTCCT GCTTAAGACC	180
	AGATGGAGTC TTCAACTAG GAGGAGGCTC GCGTTCATCT TTCAACCCCC TGCAACACAGA	240
	TTTTGCCTTC CACGCCCTGC CCTCCAGACC GCGCCACGGG GGCATAGGAT CCAGGCAGTT	300
5	TGTAGAGGAA TTTGTGCCCC CCGTCTACCT CAACCCCTAC TCGGGACCGC CGGACTCTTA	360
	TCCGGACCAG TTTATACGCC ACTACAACGT GTACAGCAAC TCTGTGAGCG GTTATAGCTG	420
	AGATTGTAAG ACTCTCCTAT CTGTCTCTGT GCTGCTTTTC CGCTTCAAGC CCCACAAGCA	480
	TGAAGGGGTT TCTGCTCCTC TTCAGCCTGC TTGTGCATTG TCCCCTAATT CATGTTGGGA	540
	CCATTAGCTT CTATGCTGCA AGGCCCGGGT CTGAGCCTAA CGCGACTTAT GTTTGTGACT	600
10	ATGGAAGCGA GTCAGATTAC AACCCACCA CGGTTCTGTG GTTGGCTCGA GAGACCGATG	660
	GCTCTGGAT CTCTGTTCTT TTCCGTCACA ACGGCTCCTC AACTGCAGCC CCCGGGGTCG	720
	TGCGCACTT TACTGACCAC AACAGCAGCA TTGTGGTGCC CCAGTATTAC CTCCTCAACA	780
	ACTCACTCTC TAAGCTCTGC TGCTCATACC GGCACAACGA GCGTTCCTAG TTTACCTGCA	840
	AACAAGCTGA CGTCCCTACC TGTACGAGC CCGGCAAGCC GCTCACCTC CGCGTCTCCC	900
15	CCGCGCTGGG AACTGCCAC CAAGCAGTCA CTTGGTTTTT TCAAAATGTA CCCATAGCTA	960
	CTGTTTACCG ACCTTGGGGC AATGTAACCT GGTTTTGTC TCCCTTCATG TGTACCTTTA	1020
	ATGTCAGCCT GAACTCCCTA CTTATTTACA ACTTTTCTGA CAAAACGGG GGGCAATACA	1080
	CAGCTCTCAT GCACTCCGGA CCGCTTCCC TCTTTCAGCT CTTAAGCCA ACGACTTGTC	1140
	TCACCAAGGT GGAGGACCCG CCGTATGCCA ACGACCCGGC CTCGCTGTG TGGCGCCAC	1200
20	TGCTTTTTCG CTTCGTCCTC TGCACGGCT GCGCGGTGTT GTTAACCGCC TTCGGTCCAT	1260
	CGATTCTATC CGGTACCGA AAGCTTATCT CAGCCCGCTT TTGGAGTCCC GAGCCCTATA	1320
	CCACCCTCCA CTAACAGTCC CCCCATGGAG CCAGACGGAG TTCATGCCGA GCAGCAGTTT	1380
	ATCCTCAATC AGATTTCCTG CGCCAACT GGCCTCCAGC GTCAAAGGGA GGAAC TAG	1438
	CTT CCC TTG TCA TGT TGC ATG CCT GTA AGC GTG GCC TCT TTT GTC CAG	1486
25	Leu Pro Leu Ser Cys Cys Met Pro Val Ser Val Ala Ser Phe Val Gln	
	1 5 10 15	
	TCA AAA CTT ACA AGC TCA GCC TCA ACG CCT CGG CCA GCG AGC ACA GCC	1534
	Ser Lys Leu Thr Ser Ser Ala Ser Thr Pro Arg Pro Ala Ser Thr Ala	
	20 25 30	
	TGC ACT TTG AAA AAA GTC CCT CCC GAT TCA CCC TGG TCA ACA CTC ACG	1582
	Cys Thr Leu Lys Lys Val Pro Pro Asp Ser Pro Trp Ser Thr Leu Thr	
	35 40 45	
30	CCG GAG CTT CTG TGC GAG TGG CCC TAC ACC ACC AGG GAG CTT CCG GCA	1630
	Pro Glu Leu Leu Cys Glu Trp Pro Tyr Thr Thr Arg Glu Leu Pro Ala	
	50 55 60	
	GCA TCC GCT GTT CCT GTT CCC ACG CCG AGT GCC TCC CCG TCC TCC TCA	1678
	Ala Ser Ala Val Pro Val Pro Thr Pro Ser Ala Ser Pro Ser Ser Ser	
	65 70 75 80	
	AGA CCC TCT GTG CCT TTA ACT TTT TAGATTAGCT GAAAGCAAAT ATAAATGGT	1732
35	Arg Pro Ser Val Pro Leu Thr Phe	
	85	
	GTGCTTACCG TAATTCTGTT TTGACTTGTC TGCTTGATTT CTCCCCCTGC GCCGTAATCC	1792
	AGTGCCCTC TTCAAACTC TCGTACCCTA TGCGATTGCG ATAGGCATAT TTTCTAAAAG	1852
	CTCTGAAGTC AACATCACTC TCAAACACTT CTCCGTTGTA GGTTACTTTC ATCTACAGAT	1912
	AAAGTCATCC ACCGGTTAAC ATCATGAAGA GAAGTGTCCT CCAGGACTTT AATCTTGTGT	1972

	ATCCGTACAA GGCTAAGAGG CCCAACATCA TGCCGCCCTT TTTTGACCGC AATGGCTTTG	2032
	TTGAAAACCA AGAAGCCACG CTAGCCATGC TTGTGGAAAA GCCGCTCAGC TTCGACAAGG	2092
	AAGGTGCGCT GACCTGGGC GTCGGACGG GCATCCGCAT TAACCCCGCG GGGCTTCTGG	2152
	AGACAAACGA CCTCGCGTCC GCTGTCTTCC CACCGCTGGC CTCCGATGAG GCCGGCAACG	2212
5	TCACGCTCAA CATGTCTGAC GGGCTATATA CTAAGGACAA CAAGCTAGCT GTCAAAGTAG	2272
	GTCCCGGGCT GTCCCTCGAC TCCAATAATG CTCTCCAGGT CCACACAGGC GACGGGCTCA	2332
	CGGTAACCGA TGACAAGGTG TCTCTAAATA CCCAAGCTCC CCTCTCGACC ACCAGCGCGG	2392
	GCCTCTCCCT ACTTCTGGGT CCCAGCTCC ACTTAGGTGA GGAGGAACGA CTAACAGTAA	2452
	ACACCGGAGC GGGCTCCAA ATTAGCAATA ACGCTCTGGC CGTAAAGTA GGTTCAGGTA	2512
10	TCACCGTAGA TGCTAAAAC CAGCTCGTG CATCCCTGGG GGACGGTCTA GAAAGCAGAG	2572
	ATAATAAAC TGTCGTTAAG GCTGGGCCG GACTTACAAT AACTAATCAA GCTCTTACTG	2632
	TTGCTACCGG GAACGGCCTT CAGGTCAACC CGGAAGGGCA ACTGCAGCTA AACATTACTG	2692
	CCGGTCAGGG CCTCAACTTT GCAAACAACA GCCTCGCCGT GGAGCTGGG TCGGGCTGCT	2752
	ATTTTCCCC TGCCAAAAC CAAGTAAGCC TTTATCCCG AGATGGAATA GACATCCGAG	2812
15	ATAATAGGGT GACTGTGCC GCTGGGCCAG GCCTGAGAAT GCTCAACCAC CAACTTGCCG	2872
	TAGCTTCCGG AGACGGTTA GAAGTCCACA GCGACCCCT CCGGTTAAAG CTCTCCACG	2932
	GCCTGACATT TGAAATGGC GCCGTACGAG CAAACTAGG ACCAGGACTT GGCACAGACG	2992
	ACTCTGGTCG GTCCGTGGT CCGACAGGTC GAGGACTTAG AGTTGCAAA GGCCTAAGTC	3052
	AGATCTTCAG CGGAAGAGGC ACCGCCATCG GCACTGATAG CAGCCTCACT CTCAACATCC	3112
20	GGGCGCCCT ACAATTTTCT GGACCCGCT TGA CTGCTAG TTTGCAAGGC AGTGGTCCGA	3172
	TTACTTACAA CAGCAACAAT GGCCTTTTCG GTCTCTCTAT AGGCCCCGGA ATGTGGGTAG	3232
	ACCAAAACAG ACTTCAGGTA AACCAGGCG CTGGTTTAGT CTTCCAAGGA AACAACCTTG	3292
	TCCAAACCT TGCGGATCCG CTGGCTATTT CCGACAGCAA AATTAGTCTC AGTCTCGGTC	3352
	CCGGCCTGAC CCAAGCTTCC AACGCCCTGA CTTTAAGTT AGGAAACGGG CTTGAATTCT	3412
25	CCAATCAAGC CGTTGCTATA AAAGCGGGCC GGGGCTTACG CTTTGAGTCT TCCTCACAAG	3472
	CTTTAGAGAG CAGCCTCACA GTCGGAATG GCTTAACGCT TACCGATACT GTGATCCGCC	3532
	CCAACCTAGG GGACGGCCTA GAGGTCAGAG ACAATAAAAT CATTGTAAAG CTGGGCGCGA	3592
	ATCTTCGTTT TGAAAACGGA GCCGTAACCG CCGCACCGT TAACCTTCT CCGCCCGAGG	3652
	CACCACCAAC TCTCACTGCA GAACCACCCC TCCGAGCCTC CAACTCCCAT CTTCAACTGT	3712
30	CCCTATCGGA GGGCTTGGTT GTGCATAACA ACGCCCTTGC TCTCCACTG GGAGACGGCA	3772
	TGGAAGTAAA TCAGCACGGA CTTACTTTAA GAGTAGGCTC GGGTTTGCAA ATGCGTGACG	3832
	GCATTTTAAC AGTTACACCC AGCGGCACTC CTATTGAGCC CAGACTGACT GCGGCACTGA	3892
	CTCAGACAGA GAATGGAATC GGGCTCGCTC TCGGCGCCGG CTTGGAATTA GACGAGAGCG	3952
	CGCTCCAAGT AAAAGTTGGG CCCGCGATGC GCCTGAACCC TGTAGAAAAG TATGTAACCC	4012
35	TGCTCCTGGG TCCTGGCCTT AGTTTTGGG AGCCGGCCAA CAGGACAAAT TATGATGTGC	4072
	GGCTTCTGT GGAGCCCCC ATGGTTTTCG GACAGCGTGG TCAGCTCACA TTTTATGTGG	4132
	GTCACGGACT ACACATTCAA AATTCCAAAC TTCAGCTCAA TTTGGGACAA GGCTCAGAA	4192
	CTGACCCCGT CACCAACCAG CTGGAAGTGC CCCTCGGTCA AGGTTTGAA ATTGAGACG	4252
	AATCCCAGGT TAGGGTTAAA TTGGGCGATG GCCTGCAGTT TGATTCACAA GCTCGCATCA	4312

-89-

CTACCGCTCC TAACATGGTC ACTGAACTC TGTGGACCGG AACAGGCAGT AATGCTAATG 4372
 TTACATGGCG GGGCTACACT GCGCCCGGCA GCAAACTCTT TTTGAGTCTC ACTCGGTTCA 4432
 GCACTGGTCT AGTTTTAGGA AACATGACTA TTGACAGCAA TGCATCCTTT GGGCAATACA 4492
 TTAACGCGGG ACACGAACAG ATCGAATGCT TTATATTGTT GGACAATCAG GGTAACTTAA 4552
 5 AAGAAGGATC TAACCTGCAA GGCACCTGGG AAGTGAAGAA CAACCCCTCT GCTTCCAAAG 4612
 CTGCTTTTTT GCCTTCACC GCCCTATACC CCATCCTCAA CGAAAGCCGA GGGAGTCTTC 4672
 CTGGAAAAAA TCTTGTGGG ATGCAAGCCA TACTGGGAGG CGGGGGCACT TGCACGTGA 4732
 TAGCCACCCT CAATGGCAGA CGCAGCAACA ACTATCCCGC GGGCCAGTCC ATAATTTTCG 4792
 TGTGGCAAGA ATTCAACACC ATAGCCCGCC AACCTCTGAA CCACTCTACA CTACTTTTT 4852
 10 CTACTGGAC TTAAATAAGT TGGAAATAAA GAGTAAACT GAATGTTTAA GTGCAACAGA 4912
 CTTTTATTGG TTTTGGCTCA CAACAAATTA CAACAGCATA GACAAGTCAT ACCGGTCAAA 4972
 CAACACAGGC TCTCGAAAAC GGGCTAACCG CTCGAAGAAT CTGTCACGCA GACGAGCAAG 5032
 TCCTAAATGT TTTTCACTC TCTTCGGGGC CAAGTTCAGC ATGTATCGGA TTTTCTGCTT 5092
 ACACCTTT 5100

15

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 88 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Leu Pro Leu Ser Cys Cys Met Pro Val Ser Val Ala Ser Phe Val Gln
 1 5 10 15
 Ser Lys Leu Thr Ser Ser Ala Ser Thr Pro Arg Pro Ala Ser Thr Ala
 20 25 30
 Cys Thr Leu Lys Lys Val Pro Pro Asp Ser Pro Trp Ser Thr Leu Thr
 35 40 45
 25 Pro Glu Leu Leu Cys Glu Trp Pro Tyr Thr Thr Arg Glu Leu Pro Ala
 50 55 60
 Ala Ser Ala Val Pro Val Pro Thr Pro Ser Ala Ser Pro Ser Ser Ser
 65 70 75 80
 Arg Pro Ser Val Pro Leu Thr Phe
 85

30

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5100 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

35

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 1915..4863

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CCTCATCAA CAACCCGTGG TGGGCACCAC CCACGTGGAA ATGCCTCGCA ACGAAGTCCT 60

-90-

	AGAACAACAT CTGACCTCAC ATGGCGCTCA AATCGCGGGC GGAGGCGCTG CGGGCGATTA	120
	CTTTAAAGC CCCACTTCAG CTCGAACCTT TATCCGCTC ACCGCCTCCT GCTTAAGACC	180
	AGATGGAGTC TTTCAACTAG GAGGAGGCTC GCGTTCATCT TTCAACCCCC TGCAACAGA	240
	TTTTGCCTTC CACGCCCTGC CCTCCAGACC GCGCCACGGG GGCATAGGAT CCAGGCAGTT	300
5	TGTAGAGGAA TTTGTGCCCC CCGTCTACCT CAACCCCTAC TCGGGACCGC CGGACTCTTA	360
	TCCGGACCAG TTTATACGCC ACTACAACGT GTACAGCAAC TCTGTGAGCG GTTATAGCTG	420
	AGATTGTAAG ACTCTCTAT CTGTCTCTGT GCTGCTTTTC CGCTTCAAGC CCCACAAGCA	480
	TGAAGGGGTT TCTGCTCATC TTCAGCCTGC TTGTGCATTG TCCCCTAATT CATGTTGGGA	540
	CCATTAGCTT CTATGCTGCA AGGCCCGGGT CTGAGCCTAA CGCGACTTAT GTTTGTGACT	600
10	ATGGAAGCGA GTCAGATTAC AACCCACCA CGTTTCTGTG GTTGGCTCGA GAGACCGATG	660
	GCTCCTGGAT CTCTGTTCTT TTCCGTCACA ACGGCTCCTC AACTGCAGCC CCCGGGGTCG	720
	TCGCGCACTT TACTGACCAC AACAGCAGCA TTGTGGTGCC CCAGTATTAC CTCCTCAACA	780
	ACTCACTCTC TAAGCTCTGC TGCTCATACC GGCACAACGA GCGTTCTCAG TTTACCTGCA	840
	AACAAGCTGA CGTCCCTACC TGTACAGAGC CCGGCAAGCC GCTCACCCTC CGCGTCTCCC	900
15	CCGCGCTGGG AACTGCCAC CAAGCAGTCA CTGTGTTTT TCAAAATGTA CCCATAGCTA	960
	CTGTTTACCG ACCTTGGGGC AATGTAACCT GGTTTTGTC TCCCTTCATG TGTACCTTTA	1020
	ATGTCAGCCT GAACCTCCCTA CTTATTTACA ACTTTTCTGA CAAAACCGGG GGGCAATACA	1080
	CAGCTCTCAT GCACTCCGGA CCTGCTTCCC TCTTTCAGCT CTTTAAGCCA ACGACTTGTG	1140
	TCACCAAGGT GGAGGACCCG CCGTATGCCA ACGACCCGGC CTCGCTGTG TGGCGCCAC	1200
20	TGCTTTTTCG CTTGCTCCTC TGCACCGGCT GCGCGGTGTT GTTAACCGCC TTCGGTCCAT	1260
	CGATTCTATC CGGTACCGGA AAGCTTATCT CAGCCCGCTT TTGGAGTCCC GAGCCCTATA	1320
	CCACCCTCCA CTAACAGTCC CCCCATGGAG CCAGACGGAG TTCATGCCGA GCAGCAGTTT	1380
	ATCCTCAATC AGATTTCTTG CGCCAACACT GCCCTCCAGC GTCAAAGGGA GGAAGTAGCT	1440
	TCCCTTGTC TGTGTCATGC CTGTAAGCGT GGCCTCTTTT GTCCAGTCAA AACTTACAAG	1500
25	CTCAGCCTCA ACGCCTCGGC CAGCGAGCAC AGCCTGCACT TTGAAAAAG TCCCTCCCGA	1560
	TTACCCCTGG TCAACACTCA CGCCGGAGCT TCTGTGCGAG TGGCCCTACA CCACCAGGGA	1620
	GCTTCCGGCA GCATCCGCTG TTCCTGTTCC CACGCCGAGT GCCTCCCCGT CCTCCTCAAG	1680
	ACCCTCTGTG CCTTAACTT TTTAGATTAG CTGAAAGCAA ATATAAATG GTGTGCTTAC	1740
	CGTAATTCTG TTTTGACTTG TGTGCTTGAT TTCTCCCCCT GCGCCGTAAT CCAGTGCCCC	1800
30	TCTTCAAAC TCTCGTACCC TATGCGATTC GCATAGGCAT ATTTTCTAAA AGCTCTGAAG	1860
	TCAACATCAC TCTCAAACAC TTCTCCGTTG TAGGTTACTT TCATCTACAG ATAA AGT	1917
		Ser 1
	CAT CCA CCG GTT AAC ATC ATG AAG AGA AGT GTG CCC CAG GAC TTT AAT	1965
	His Pro Pro Val Asn Ile Met Lys Arg Ser Val Pro Gln Asp Phe Asn	
	5 10 15	
35	CTT GTG TAT CCG TAC AAG GCT AAG AGG CCC AAC ATC ATG CCG CCC TTT	2013
	Leu Val Tyr Pro Tyr Lys Ala Lys Arg Pro Asn Ile Met Pro Pro Phe	
	20 25 30	
	TTT GAC CGC AAT GGC TTT GTT GAA AAC CAA GAA GCC ACG CTA GCC ATG	2061
	Phe Asp Arg Asn Gly Phe Val Glu Asn Gln Glu Ala Thr Leu Ala Met	
	35 40 45	
	CTT GTG GAA AAG CCG CTC ACG TTC GAC AAG GAA GGT GCG CTG ACC CTG	2109

-91-

	Leu Val Glu Lys Pro Leu Thr Phe Asp Lys Glu Gly Ala Leu Thr Leu	
	50 55 60 65	
	GGC GTC GGA CGC GGC ATC CGC ATT AAC CCC GCG GGG CTT CTG GAG ACA	2157
	Gly Val Gly Arg Gly Ile Arg Ile Asn Pro Ala Gly Leu Leu Glu Thr	
	70 75 80	
5	AAC GAC CTC GCG TCC GCT GTC TTC CCA CCG CTG GCC TCC GAT GAG GCC	2205
	Asn Asp Leu Ala Ser Ala Val Phe Pro Pro Leu Ala Ser Asp Glu Ala	
	85 90 95	
	GGC AAC GTC ACG CTC AAC ATG TCT GAC GGG CTA TAT ACT AAG GAC AAC	2253
	Gly Asn Val Thr Leu Asn Met Ser Asp Gly Leu Tyr Thr Lys Asp Asn	
	100 105 110	
	AAG CTA GCT GTC AAA GTA GGT CCC GGG CTG TCC CTC GAC TCC AAT AAT	2301
	Lys Leu Ala Val Lys Val Gly Pro Gly Leu Ser Leu Asp Ser Asn Asn	
	115 120 125	
10	GCT CTC CAG GTC CAC ACA GGC GAC GGG CTC ACG GTA ACC GAT GAC AAG	2349
	Ala Leu Gln Val His Thr Gly Asp Gly Leu Thr Val Thr Asp Asp Lys	
	130 135 140 145	
	GTG TCT CTA AAT ACC CAA GCT CCC CTC TCG ACC ACC AGC GCG GGC CTC	2397
	Val Ser Leu Asn Thr Gln Ala Pro Leu Ser Thr Thr Ser Ala Gly Leu	
	150 155 160	
15	TCC CTA CTT CTG GGT CCC AGC CTC CAC TTA GGT GAG GAG GAA CGA CTA	2445
	Ser Leu Leu Gly Pro Ser Leu His Leu Gly Glu Glu Glu Arg Leu	
	165 170 175	
	ACA GTA AAC ACC GGA GCG GGC CTC CAA ATT AGC AAT AAC GCT CTG GCC	2493
	Thr Val Asn Thr Gly Ala Gly Leu Gln Ile Ser Asn Asn Ala Leu Ala	
	180 185 190	
	GTA AAA GTA GGT TCA GGT ATC ACC GTA GAT GCT CAA AAC CAG CTC GCT	2541
	Val Lys Val Gly Ser Gly Ile Thr Val Asp Ala Gln Asn Gln Leu Ala	
	195 200 205	
20	GCA TCC CTG GGG GAC GGT CTA GAA AGC AGA GAT AAT AAA ACT GTC GTT	2589
	Ala Ser Leu Gly Asp Gly Leu Glu Ser Arg Asp Asn Lys Thr Val Val	
	210 215 220 225	
	AAG GCT GGG CCC GGA CTT ACA ATA ACT AAT CAA GCT CTT ACT GTT GCT	2637
	Lys Ala Gly Pro Gly Leu Thr Ile Thr Asn Gln Ala Leu Thr Val Ala	
	230 235 240	
25	ACC GGG AAC GGC CTT CAG GTC AAC CCG GAA GGG CAA CTG CAG CTA AAC	2685
	Thr Gly Asn Gly Leu Gln Val Asn Pro Glu Gly Gln Leu Gln Leu Asn	
	245 250 255	
	ATT ACT GCC GGT CAG GGC CTC AAC TTT GCA AAC AAC AGC CTC GCC GTG	2733
	Ile Thr Ala Gly Gln Gly Leu Asn Phe Ala Asn Asn Ser Leu Ala Val	
	260 265 270	
	GAG CTG GGC TCG GGC CTG CAT TTT CCC CCT GGC CAA AAC CAA GTA AGC	2781
	Glu Leu Gly Ser Gly Leu His Phe Pro Pro Gly Gln Asn Gln Val Ser	
	275 280 285	
30	CTT TAT CCC GGA GAT GGA ATA GAC ATC CGA GAT AAT AGG GTG ACT GTG	2829
	Leu Tyr Pro Gly Asp Gly Ile Asp Ile Arg Asp Asn Arg Val Thr Val	
	290 295 300 305	
	CCC GCT GGG CCA GGC CTG AGA ATG CTC AAC CAC CAA CTT GCC GTA GCT	2877
	Pro Ala Gly Pro Gly Leu Arg Met Leu Asn His Gln Leu Ala Val Ala	
	310 315 320	
35	TCC GGA GAC GGT TTA GAA GTC CAC AGC GAC ACC CTC CGG TTA AAG CTC	2925
	Ser Gly Asp Gly Leu Glu Val His Ser Asp Thr Leu Arg Leu Lys Leu	
	325 330 335	
	TCC CAC GGC CTG ACA TTT GAA AAT GGC GCC GTA CGA GCA AAA CTA GGA	2973
	Ser His Gly Leu Thr Phe Glu Asn Gly Ala Val Arg Ala Lys Leu Gly	
	340 345 350	
	CCA GGA CTT GGC ACA GAC GAC TCT GGT CGG TCC GTG GTT CGC ACA GGT	3021
	Pro Gly Leu Gly Thr Asp Asp Ser Gly Arg Ser Val Val Arg Thr Gly	

-92-

	355	360	365	
	CGA GGA CTT AGA GTT GCA AAC GGC CAA GTC CAG ATC TTC AGC GGA AGA Arg Gly Leu Arg Val Ala Asn Gly Gln Val Gln Ile Phe Ser Gly Arg 370 375 380 385	3069		
5	GGC ACC GCC ATC GGC ACT GAT AGC AGC CTC ACT CTC AAC ATC CGG GCG Gly Thr Ala Ile Gly Thr Asp Ser Ser Leu Thr Leu Asn Ile Arg Ala 390 395 400	3117		
	CCC CTA CAA TTT TCT GGA CCC GCC TTG ACT GCT AGT TTG CAA GGC AGT Pro Leu Gln Phe Ser Gly Pro Ala Leu Thr Ala Ser Leu Gln Gly Ser 405 410 415	3165		
	GGT CCG ATT ACT TAC AAC AGC AAC AAT GGC ACT TTC GGT CTC TCT ATA Gly Pro Ile Thr Tyr Asn Ser Asn Asn Gly Thr Phe Gly Leu Ser Ile 420 425 430	3213		
10	GGC CCC GGA ATG TGG GTA GAC CAA AAC AGA CTT CAG GTA AAC CCA GGC Gly Pro Gly Met Trp Val Asp Gln Asn Arg Leu Gln Val Asn Pro Gly 435 440 445	3261		
	GCT GGT TTA GTC TTC CAA GGA AAC AAC CTT GTC CCA AAC CTT GCG GAT Ala Gly Leu Val Phe Gln Gly Asn Asn Leu Val Pro Asn Leu Ala Asp 450 455 460 465	3309		
15	CCG CTG GCT ATT TCC GAC AGC AAA ATT AGT CTC AGT CTC GGT CCC GGC Pro Leu Ala Ile Ser Asp Ser Lys Ile Ser Leu Ser Leu Gly Pro Gly 470 475 480	3357		
	CTG ACC CAA GCT TCC AAC GCC CTG ACT TTA AGT TTA GGA AAC GGG CTT Leu Thr Gln Ala Ser Asn Ala Leu Thr Leu Ser Leu Gly Asn Gly Leu 485 490 495	3405		
	GAA TTC TCC AAT CAA GCC GTT GCT ATA AAA GCG GGC CGG GGC TTA CGC Glu Phe Ser Asn Gln Ala Val Ala Ile Lys Ala Gly Arg Gly Leu Arg 500 505 510	3453		
20	TTT GAG TCT TCC TCA CAA GCT TTA GAG AGC AGC CTC ACA GTC GGA AAT Phe Glu Ser Ser Ser Gln Ala Leu Glu Ser Ser Leu Thr Val Gly Asn 515 520 525	3501		
	GGC TTA ACG CTT ACC GAT ACT GTG ATC CGC CCC AAC CTA GGG GAC GGC Gly Leu Thr Leu Thr Asp Thr Val Ile Arg Pro Asn Leu Gly Asp Gly 530 535 540 545	3549		
25	CTA GAG GTC AGA GAC AAT AAA ATC ATT GTT AAG CTG GGC GCG AAT CTT Leu Glu Val Arg Asp Asn Lys Ile Ile Val Lys Leu Gly Ala Asn Leu 550 555 560	3597		
	CGT TTT GAA AAC GGA GCC GTA ACC GCC GGC ACC GTT AAC CCT TCT GCG Arg Phe Glu Asn Gly Ala Val Thr Ala Gly Thr Val Asn Pro Ser Ala 565 570 575	3645		
	CCC GAG GCA CCA CCA ACT CTC ACT GCA GAA CCA CCC CTC CGA GCC TCC Pro Glu Ala Pro Pro Thr Leu Thr Ala Glu Pro Pro Leu Arg Ala Ser 580 585 590	3693		
30	AAC TCC CAT CTT CAA CTG TCC CTA TCG GAG GGC TTG GTT GTG CAT AAC Asn Ser His Leu Gln Leu Ser Leu Ser Glu Gly Leu Val Val His Asn 595 600 605	3741		
	AAC GCC CTT GCT CTC CAA CTG GGA GAC GGC ATG GAA GTA AAT CAG CAC Asn Ala Leu Ala Leu Gln Leu Gly Asp Gly Met Glu Val Asn Gln His 610 615 620 625	3789		
35	GGA CTT ACT TTA AGA GTA GGC TCG GGT TTG CAA ATG CGT GAC GGC ATT Gly Leu Thr Leu Arg Val Gly Ser Gly Leu Gln Met Arg Asp Gly Ile 630 635 640	3837		
	TTA ACA GTT ACA CCC AGC GGC ACT CCT ATT GAG CCC AGA CTG ACT GCC Leu Thr Val Thr Pro Ser Gly Thr Pro Ile Glu Pro Arg Leu Thr Ala 645 650 655	3885		
	CCA CTG ACT CAG ACA GAG AAT GGA ATC GGG CTC GCT CTC GGC GCC GGC Pro Leu Thr Gln Thr Glu Asn Gly Ile Gly Leu Ala Leu Gly Ala Gly 660 665 670	3933		

-93-

	TTG GAA TTA GAC GAG AGC GCG CTC CAA GTA AAA GTT GGG CCC GGC ATG Leu Glu Leu Asp Glu Ser Ala Leu Gln Val Lys Val Gly Pro Gly Met 675 680 685	3981
	CGC CTG AAC CCT GTA GAA AAG TAT GTA ACC CTG CTC CTG GGT CCT GGC Arg Leu Asn Pro Val Glu Lys Tyr Val Thr Leu Leu Leu Gly Pro Gly 690 695 700 705	4029
5	CTT AGT TTT GGG CAG CCG GCC AAC AGG ACA AAT TAT GAT GTG CGC GTT Leu Ser Phe Gly Gln Pro Ala Asn Arg Thr Asn Tyr Asp Val Arg Val 710 715 720	4077
	TCT GTG GAG CCC CCC ATG GTT TTC GGA CAG CGT GGT CAG CTC ACA TTT Ser Val Glu Pro Pro Met Val Phe Gly Gln Arg Gly Gln Leu Thr Phe 725 730 735	4125
10	TTA GTG GGT CAC GGA CTA CAC ATT CAA AAT TCC AAA CTT CAG CTC AAT Leu Val Gly His Gly Leu His Ile Gln Asn Ser Lys Leu Gln Leu Asn 740 745 750	4173
	TTG GGA CAA GGC CTC AGA ACT GAC CCC GTC ACC AAC CAG CTG GAA GTG Leu Gly Gln Gly Leu Arg Thr Asp Pro Val Thr Asn Gln Leu Glu Val 755 760 765	4221
	CCC CTC GGT CAA GGT TTG GAA ATT GCA GAC GAA TCC CAG GTT AGG GTT Pro Leu Gly Gln Gly Leu Glu Ile Ala Asp Glu Ser Gln Val Arg Val 770 775 780 785	4269
15	AAA TTG GGC GAT GGC CTG CAG TTT GAT TCA CAA GCT CGC ATC ACT ACC Lys Leu Gly Asp Gly Leu Gln Phe Asp Ser Gln Ala Arg Ile Thr Thr 790 795 800	4317
	GCT CCT AAC ATG GTC ACT GAA ACT CTG TGG ACC GGA ACA GGC AGT AAT Ala Pro Asn Met Val Thr Glu Thr Leu Trp Thr Gly Thr Gly Ser Asn 805 810 815	4365
20	GCT AAT GTT ACA TGG CGG GGC TAC ACT GCC CCC GGC AGC AAA CTC TTT Ala Asn Val Thr Trp Arg Gly Tyr Thr Ala Pro Gly Ser Lys Leu Phe 820 825 830	4413
	TTG AGT CTC ACT CGG TTC AGC ACT GGT CTA GTT TTA GGA AAC ATG ACT Leu Ser Leu Thr Arg Phe Ser Thr Gly Leu Val Leu Gly Asn Met Thr 835 840 845	4461
	ATT GAC AGC AAT GCA TCC TTT GGG CAA TAC ATT AAC GCG GGA CAC GAA Ile Asp Ser Asn Ala Ser Phe Gly Gln Tyr Ile Asn Ala Gly His Glu 850 855 860 865	4509
25	CAG ATC GAA TGC TTT ATA TTG TTG GAC AAT CAG GGT AAC CTA AAA GAA Gln Ile Glu Cys Phe Ile Leu Leu Asp Asn Gln Gly Asn Leu Lys Glu 870 875 880	4557
	GGA TCT AAC TTG CAA GGC ACT TGG GAA GTG AAG AAC AAC CCC TCT GCT Gly Ser Asn Leu Gln Gly Thr Trp Glu Val Lys Asn Asn Pro Ser Ala 885 890 895	4605
30	TCC AAA GCT GCT TTT TTG CCT TCC ACC GCC CTA TAC CCC ATC CTC AAC Ser Lys Ala Ala Phe Leu Pro Ser Thr Ala Leu Tyr Pro Ile Leu Asn 900 905 910	4653
	GAA AGC CGA GGG AGT CTT CCT GGA AAA AAT CTT GTG GGC ATG CAA GCC Glu Ser Arg Gly Ser Leu Pro Gly Lys Asn Leu Val Gly Met Gln Ala 915 920 925	4701
	ATA CTG GGA GGC GGG GGC ACT TGC ACT GTG ATA GCC ACC CTC AAT GGC Ile Leu Gly Gly Gly Gly Thr Cys Thr Val Ile Ala Thr Leu Asn Gly 930 935 940 945	4749
35	AGA CGC AGC AAC AAC TAT CCC GCG GGC CAG TCC ATA ATT TTC GTG TGG Arg Arg Ser Asn Asn Tyr Pro Ala Gly Gln Ser Ile Ile Phe Val Trp 950 955 960	4797
	CAA GAA TTC AAC ACC ATA GCC CGC CAA CCT CTG AAC CAC TCT ACA CTT Gln Glu Phe Asn Thr Ile Ala Arg Gln Pro Leu Asn His Ser Thr Leu 965 970 975	4845
	ACT TTT TCT TAC TGG ACT TAAATAAGTT GGAAATAAAG AGTTAACTG	4893

-94-

Thr Phe Ser Tyr Trp Thr
980

AATGTTTAAG TGCAACAGAC TTTTATTGGT TTTGGCTCAC AACAAATTAC AACAGCATAG 4953
ACAAGTCATA CCGGTCAAAC AACACAGGCT CTCGAAAACG GGCTAACCGC TCCAAGAATC 5013
TGTCAACGAG ACGAGCAAGT CCTAAATGTT TTTTCACTCT CTTCGGGGCC AAGTTCAGCA 5073
TGTATCGGAT TTTCTGCTTA CACCTTT 5100

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 983 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Ser His Pro Pro Val Asn Ile Met Lys Arg Ser Val Pro Gln Asp Phe
1 5 10 15
Asn Leu Val Tyr Pro Tyr Lys Ala Lys Arg Pro Asn Ile Met Pro Pro
20 25 30
Phe Phe Asp Arg Asn Gly Phe Val Glu Asn Gln Glu Ala Thr Leu Ala
35 40 45
Met Leu Val Glu Lys Pro Leu Thr Phe Asp Lys Glu Gly Ala Leu Thr
50 55 60
Leu Gly Val Gly Arg Gly Ile Arg Ile Asn Pro Ala Gly Leu Leu Glu
65 70 75 80
Thr Asn Asp Leu Ala Ser Ala Val Phe Pro Pro Leu Ala Ser Asp Glu
85 90 95
Ala Gly Asn Val Thr Leu Asn Met Ser Asp Gly Leu Tyr Thr Lys Asp
100 105 110
Asn Lys Leu Ala Val Lys Val Gly Pro Gly Leu Ser Leu Asp Ser Asn
115 120 125
Asn Ala Leu Gln Val His Thr Gly Asp Gly Leu Thr Val Thr Asp Asp
130 135 140
Lys Val Ser Leu Asn Thr Gln Ala Pro Leu Ser Thr Thr Ser Ala Gly
145 150 155 160
Leu Ser Leu Leu Leu Gly Pro Ser Leu His Leu Gly Glu Glu Glu Arg
165 170 175
Leu Thr Val Asn Thr Gly Ala Gly Leu Gln Ile Ser Asn Asn Ala Leu
180 185 190
Ala Val Lys Val Gly Ser Gly Ile Thr Val Asp Ala Gln Asn Gln Leu
195 200 205
Ala Ala Ser Leu Gly Asp Gly Leu Glu Ser Arg Asp Asn Lys Thr Val
210 215 220
Val Lys Ala Gly Pro Gly Leu Thr Ile Thr Asn Gln Ala Leu Thr Val
225 230 235 240
Ala Thr Gly Asn Gly Leu Gln Val Asn Pro Glu Gly Gln Leu Gln Leu
245 250 255
Asn Ile Thr Ala Gly Gln Gly Leu Asn Phe Ala Asn Asn Ser Leu Ala
260 265 270
Val Glu Leu Gly Ser Gly Leu His Phe Pro Pro Gly Gln Asn Gln Val
275 280 285

-95-

Ser Leu Tyr Pro Gly Asp Gly Ile Asp Ile Arg Asp Asn Arg Val Thr
 290 295 300
 Val Pro Ala Gly Pro Gly Leu Arg Met Leu Asn His Gln Leu Ala Val
 305 310 315 320
 Ala Ser Gly Asp Gly Leu Glu Val His Ser Asp Thr Leu Arg Leu Lys
 325 330 335
 5 Leu Ser His Gly Leu Thr Phe Glu Asn Gly Ala Val Arg Ala Lys Leu
 340 345 350
 Gly Pro Gly Leu Gly Thr Asp Asp Ser Gly Arg Ser Val Val Arg Thr
 355 360 365
 Gly Arg Gly Leu Arg Val Ala Asn Gly Gln Val Gln Ile Phe Ser Gly
 370 375 380
 10 Arg Gly Thr Ala Ile Gly Thr Asp Ser Ser Leu Thr Leu Asn Ile Arg
 385 390 395 400
 Ala Pro Leu Gln Phe Ser Gly Pro Ala Leu Thr Ala Ser Leu Gln Gly
 405 410 415
 Ser Gly Pro Ile Thr Tyr Asn Ser Asn Asn Gly Thr Phe Gly Leu Ser
 420 425 430
 15 Ile Gly Pro Gly Met Trp Val Asp Gln Asn Arg Leu Gln Val Asn Pro
 435 440 445
 Gly Ala Gly Leu Val Phe Gln Gly Asn Asn Leu Val Pro Asn Leu Ala
 450 455 460
 Asp Pro Leu Ala Ile Ser Asp Ser Lys Ile Ser Leu Ser Leu Gly Pro
 465 470 475 480
 Gly Leu Thr Gln Ala Ser Asn Ala Leu Thr Leu Ser Leu Gly Asn Gly
 485 490 495
 20 Leu Glu Phe Ser Asn Gln Ala Val Ala Ile Lys Ala Gly Arg Gly Leu
 500 505 510
 Arg Phe Glu Ser Ser Ser Gln Ala Leu Glu Ser Ser Leu Thr Val Gly
 515 520 525
 Asn Gly Leu Thr Leu Thr Asp Thr Val Ile Arg Pro Asn Leu Gly Asp
 530 535 540
 25 Gly Leu Glu Val Arg Asp Asn Lys Ile Ile Val Lys Leu Gly Ala Asn
 545 550 555 560
 Leu Arg Phe Glu Asn Gly Ala Val Thr Ala Gly Thr Val Asn Pro Ser
 565 570 575
 Ala Pro Glu Ala Pro Pro Thr Leu Thr Ala Glu Pro Pro Leu Arg Ala
 580 585 590
 30 Ser Asn Ser His Leu Gln Leu Ser Leu Ser Glu Gly Leu Val Val His
 595 600 605
 Asn Asn Ala Leu Ala Leu Gln Leu Gly Asp Gly Met Glu Val Asn Gln
 610 615 620
 His Gly Leu Thr Leu Arg Val Gly Ser Gly Leu Gln Met Arg Asp Gly
 625 630 635 640
 Ile Leu Thr Val Thr Pro Ser Gly Thr Pro Ile Glu Pro Arg Leu Thr
 645 650 655
 35 Ala Pro Leu Thr Gln Thr Glu Asn Gly Ile Gly Leu Ala Leu Gly Ala
 660 665 670
 Gly Leu Glu Leu Asp Glu Ser Ala Leu Gln Val Lys Val Gly Pro Gly
 675 680 685
 Met Arg Leu Asn Pro Val Glu Lys Tyr Val Thr Leu Leu Leu Gly Pro
 690 695 700

-96-

Gly Leu Ser Phe Gly Gln Pro Ala Asn Arg Thr Asn Tyr Asp Val Arg
 705 710 715 720
 Val Ser Val Glu Pro Pro Met Val Phe Gly Gln Arg Gly Gln Leu Thr
 725 730 735
 Phe Leu Val Gly His Gly Leu His Ile Gln Asn Ser Lys Leu Gln Leu
 740 745 750
 5 Asn Leu Gly Gln Gly Leu Arg Thr Asp Pro Val Thr Asn Gln Leu Glu
 755 760 765
 Val Pro Leu Gly Gln Gly Leu Glu Ile Ala Asp Glu Ser Gln Val Arg
 770 775 780
 Val Lys Leu Gly Asp Gly Leu Gln Phe Asp Ser Gln Ala Arg Ile Thr
 785 790 795 800
 10 Thr Ala Pro Asn Met Val Thr Glu Thr Leu Trp Thr Gly Thr Gly Ser
 805 810 815
 Asn Ala Asn Val Thr Trp Arg Gly Tyr Thr Ala Pro Gly Ser Lys Leu
 820 825 830
 Phe Leu Ser Leu Thr Arg Phe Ser Thr Gly Leu Val Leu Gly Asn Met
 835 840 845
 15 Thr Ile Asp Ser Asn Ala Ser Phe Gly Gln Tyr Ile Asn Ala Gly His
 850 855 860
 Glu Gln Ile Glu Cys Phe Ile Leu Leu Asp Asn Gln Gly Asn Leu Lys
 865 870 875 880
 Glu Gly Ser Asn Leu Gln Gly Thr Trp Glu Val Lys Asn Asn Pro Ser
 885 890 895
 Ala Ser Lys Ala Ala Phe Leu Pro Ser Thr Ala Leu Tyr Pro Ile Leu
 900 905 910
 20 Asn Glu Ser Arg Gly Ser Leu Pro Gly Lys Asn Leu Val Gly Met Gln
 915 920 925
 Ala Ile Leu Gly Gly Gly Gly Thr Cys Thr Val Ile Ala Thr Leu Asn
 930 935 940
 Gly Arg Arg Ser Asn Asn Tyr Pro Ala Gly Gln Ser Ile Ile Phe Val
 945 950 955 960
 25 Trp Gln Glu Phe Asn Thr Ile Ala Arg Gln Pro Leu Asn His Ser Thr
 965 970 975
 Leu Thr Phe Ser Tyr Trp Thr
 980

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 227 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

35 Met Ser Lys Glu Ile Pro Thr Pro Tyr Met Trp Ser Tyr Gln Pro Gln
 1 5 10 15
 Met Gly Leu Ala Ala Gly Ala Ala Gln Asp Tyr Ser Thr Arg Ile Asn
 20 25 30
 Tyr Met Ser Ala Gly Pro His Met Ile Ser Arg Val Asn Gly Ile Arg
 35 40 45

-97-

Ala His Arg Asn Arg Ile Leu Leu Glu Gln Ala Ala Ile Thr Thr Thr
 50 55 60
 Pro Arg Asn Asn Leu Asn Pro Arg Ser Trp Pro Ala Ala Leu Val Tyr
 65 70 75 80
 Gln Glu Ser Pro Ala Pro Thr Thr Val Val Leu Pro Arg Asp Ala Gln
 85 90 95
 Ala Glu Val Gln Met Thr Asn Ser Gly Ala Gln Leu Ala Gly Gly Phe
 100 105 110
 Arg His Arg Val Arg Ser Pro Gly Gln Gly Ile Thr His Leu Lys Ile
 115 120 125
 Arg Gly Arg Gly Ile Gln Leu Asn Asp Glu Ser Val Ser Ser Ser Leu
 130 135 140
 Gly Leu Arg Pro Asp Gly Thr Phe Gln Ile Gly Gly Ala Gly Arg Ser
 145 150 155 160
 Ser Phe Thr Pro Arg Gln Ala Ile Leu Thr Leu Gln Thr Ser Ser Ser
 165 170 175
 Glu Pro Arg Ser Gly Gly Ile Gly Thr Leu Gln Phe Ile Glu Glu Phe
 180 185 190
 Val Pro Ser Val Tyr Phe Asn Pro Phe Ser Gly Pro Pro Gly His Tyr
 195 200 205
 Pro Asp Gln Phe Ile Pro Asn Phe Asp Ala Val Lys Asp Ser Ala Asp
 210 215 220
 Gly Tyr Asp
 225

(2) INFORMATION FOR SEQ ID NO:28:

- 20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 128 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: protein

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Met Thr Asp Thr Leu Asp Leu Glu Met Asp Gly Ile Ile Thr Glu Gln
 1 5 10 15
 Arg Leu Leu Glu Arg Arg Arg Ala Ala Ala Glu Gln Gln Arg Met Asn
 20 25 30
 Gln Glu Leu Gln Asp Met Val Asn Leu His Gln Cys Lys Arg Gly Ile
 35 40 45
 Phe Cys Leu Val Lys Gln Ala Lys Val Thr Tyr Asp Ser Asn Thr Thr
 50 55 60
 Gly His Arg Leu Ser Tyr Lys Leu Pro Thr Lys Arg Gln Lys Leu Val
 65 70 75 80
 Val Met Val Gly Glu Lys Pro Ile Thr Ile Thr Gln His Ser Val Glu
 85 90 95
 Thr Glu Gly Cys Ile His Ser Pro Cys Gln Gly Pro Glu Asp Leu Cys
 100 105 110
 Thr Leu Ile Lys Thr Leu Cys Gly Leu Lys Asp Leu Ile Pro Phe Asn
 115 120 125

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 582 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Met Lys Arg Ala Arg Pro Ser Glu Asp Thr Phe Asn Pro Val Tyr Pro
 1 5 10 15

Tyr Asp Thr Glu Thr Gly Pro Pro Thr Val Pro Phe Leu Thr Pro Pro
 20 25 30

10 Phe Val Ser Pro Asn Gly Phe Gln Glu Ser Pro Pro Gly Val Leu Ser
 35 40 45

Leu Arg Val Ser Glu Pro Leu Asp Thr Ser His Gly Met Leu Ala Leu
 50 55 60

Lys Met Gly Ser Gly Leu Thr Leu Asp Lys Ala Gly Asn Leu Thr Ser
 65 70 75 80

15 Gln Asn Val Thr Thr Val Thr Gln Pro Leu Lys Lys Thr Lys Ser Asn
 85 90 95

Ile Ser Leu Asp Thr Ser Ala Pro Leu Thr Ile Thr Ser Gly Ala Leu
 100 105 110

Thr Val Ala Thr Thr Ala Pro Leu Ile Val Thr Ser Gly Ala Leu Ser
 115 120 125

Val Gln Ser Gln Ala Pro Leu Thr Val Gln Asp Ser Lys Leu Ser Ile
 130 135 140

20 Ala Thr Lys Gly Pro Ile Thr Val Ser Asp Gly Lys Leu Ala Leu Gln
 145 150 155 160

Thr Ser Ala Pro Leu Ser Gly Ser Asp Ser Asp Thr Leu Thr Val Thr
 165 170 175

Ala Ser Pro Pro Leu Thr Thr Ala Thr Gly Ser Leu Gly Ile Asn Met
 180 185 190

25 Glu Asp Pro Ile Tyr Val Asn Asn Gly Lys Ile Gly Ile Lys Ile Ser
 195 200 205

Gly Pro Leu Gln Val Ala Gln Asn Ser Asp Thr Leu Thr Val Val Thr
 210 215 220

Gly Pro Gly Val Thr Val Glu Gln Asn Ser Leu Arg Thr Lys Val Ala
 225 230 235 240

30 Gly Ala Ile Gly Tyr Asp Ser Ser Asn Asn Met Glu Ile Lys Thr Gly
 245 250 255

Gly Gly Met Arg Ile Asn Asn Asn Leu Leu Ile Leu Asp Val Asp Tyr
 260 265 270

Pro Phe Asp Ala Gln Thr Lys Leu Arg Leu Lys Leu Gly Gln Gly Pro
 275 280 285

Leu Tyr Ile Asn Ala Ser His Asn Leu Asp Ile Asn Tyr Asn Arg Gly
 290 295 300

35 Leu Tyr Leu Phe Asn Ala Ser Asn Asn Thr Lys Lys Leu Glu Val Ser
 305 310 315 320

Ile Lys Lys Ser Ser Gly Leu Asn Phe Asp Asn Thr Ala Ile Ala Ile
 325 330 335

Asn Ala Gly Lys Gly Leu Glu Phe Asp Thr Asn Thr Ser Glu Ser Pro
 340 345 350

-99-

Asp Ile Asn Pro Ile Lys Thr Lys Ile Gly Ser Gly Ile Asp Tyr Asn
 355 360 365
 Glu Asn Gly Ala Met Ile Thr Lys Leu Gly Ala Gly Leu Ser Phe Asp
 370 375 380
 Asn Ser Gly Ala Ile Thr Ile Gly Asn Lys Asn Asp Asp Lys Leu Thr
 385 390 395 400
 5 Leu Trp Thr Thr Pro Asp Pro Ser Pro Asn Cys Arg Ile His Ser Asp
 405 410 415
 Asn Asp Cys Lys Phe Thr Leu Val Leu Thr Lys Cys Gly Ser Gln Val
 420 425 430
 Leu Ala Thr Val Ala Ala Leu Ala Val Ser Gly Asp Leu Ser Ser Met
 435 440 445
 10 Thr Gly Thr Val Ala Ser Val Ser Ile Phe Leu Arg Phe Asp Gln Asn
 450 455 460
 Gly Val Leu Met Glu Asn Ser Ser Leu Lys Lys His Tyr Trp Asn Phe
 465 470 475 480
 Arg Asn Gly Asn Ser Thr Asn Ala Asn Pro Tyr Thr Asn Ala Val Gly
 485 490 495
 15 Phe Met Pro Asn Leu Leu Ala Tyr Pro Lys Thr Gln Ser Gln Thr Ala
 500 505 510
 Lys Asn Asn Ile Val Ser Gln Val Tyr Leu His Gly Asp Lys Thr Lys
 515 520 525
 Pro Met Ile Leu Thr Ile Thr Leu Asn Gly Thr Ser Glu Ser Thr Glu
 530 535 540
 20 Thr Ser Glu Val Ser Thr Tyr Ser Met Ser Phe Thr Trp Ser Trp Glu
 545 550 555 560
 Ser Gly Lys Tyr Thr Thr Glu Thr Phe Ala Thr Asn Ser Tyr Thr Phe
 565 570 575
 Ser Tyr Ile Ala Gln Glu
 580

(2) INFORMATION FOR SEQ ID NO:30:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 7 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: peptide
 (ix) FEATURE:
 30 (A) NAME/KEY: Modified-site
 (B) LOCATION: 2
 (D) OTHER INFORMATION: /note= "This position is X2."
 (ix) FEATURE:
 (A) NAME/KEY: Modified-site
 (B) LOCATION: 4
 (D) OTHER INFORMATION: /note= "This position is X13."
 35 (ix) FEATURE:
 (A) NAME/KEY: Modified-site
 (B) LOCATION: 6
 (D) OTHER INFORMATION: /note= "This position is X2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Cys Xaa Cys Xaa Cys Xaa Cys
 1 5

-100-

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 7 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Gln Ser Ser Xaa Ser Thr Ser
 1 5

(2) INFORMATION FOR SEQ ID NO:32:

10

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Pro Leu Leu Phe Ala Phe Val Leu Cys Thr Gly Cys Ala Val Leu Leu
 1 5 10 15

Thr Ala Phe Gly Pro Ser Ile Leu Ser Gly Thr
 20 25

(2) INFORMATION FOR SEQ ID NO:33:

20

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 57 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Glu Glu Val Thr Ser His Phe Phe Leu Asp Cys Pro Glu Asp Pro Ser
 1 5 10 15

Arg Glu Cys Ser Ser Cys Gly Phe His Gln Ala Gln Ser Gly Ile Pro
 20 25 30

Gly Ile Met Cys Ser Leu Cys Tyr Met Arg Gln Thr Tyr His Cys Ile
 35 40 45

30

Tyr Ser Pro Val Ser Glu Glu Glu Met
 50 55

(2) INFORMATION FOR SEQ ID NO:34:

35

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 12 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Val Asp Leu Glu Cys His Glu Val Leu Pro Pro Ser
 1 5 10

-101-

Claims

1. A live recombinant bovine adenovirus vector (BAV) system selected from the group consisting of:
- 5 of:
- (a) a system wherein part or all of the E1 gene region is replaced by a heterologous nucleotide sequence encoding a foreign gene or fragment thereof;
- 10 (b) a system wherein a part or all of the E3 gene region is replaced by a heterologous nucleotide sequence encoding a foreign gene or fragment thereof; and
- (c) a system wherein part or all of the E1 gene region and part or all of the E3 gene region are deleted and a heterologous nucleotide sequence encoding a foreign gene or fragment thereof is inserted into at least one of the deletions.
- 15
2. The BAV system of claim 1 which is a bovine adenovirus type 3.
- 20
3. The BAV system of claim 1 wherein (a) a recombinant BAV wherein part or all of the E1 gene region is replaced by a heterologous nucleotide sequence encoding a foreign gene or fragment thereof.
- 25
4. The BAV system of claim 1 wherein (b) a recombinant BAV wherein a part or all of the E3 gene region is replaced by a heterologous nucleotide sequence encoding a foreign gene or fragment thereof.
- 30
5. The BAV system of claim 1 wherein the foreign nucleotide sequence is with or without the control of an exogenous promoter.
- 35
6. The BAV system of claim 1 wherein (c) a system wherein part or all of the E1 gene region and part or all of the E3 gene region are deleted and a

-102-

heterologous nucleotide sequence encoding a foreign gene or fragment thereof is inserted into at least one of the deletions.

5 7. A recombinant vector system comprising the entire BAV genome and a plasmid capable of generating a recombinant virus by in vivo recombination following cotransfection of a suitable cell line comprising the entire BAV genome
10 representing the wild-type BAV genome and a plasmid comprising an adenovirus left end nucleotide sequences containing the E1A gene region or a plasmid comprising adenovirus right end sequences containing the E3 gene region, the plasmid with a heterologous nucleotide
15 sequence encoding a foreign gene or fragment thereof substituted for part or all of the E1 and/or E3 gene regions, respectively.

8. A recombinant bovine adenovirus vector
20 system comprising two plasmids capable of generating a recombinant virus by in vivo recombination following cotransfection of a cell line comprising

(1) a first plasmid comprising the entire BAV genome except for a deletion of part or all
25 of the E1 and/or E3 gene regions, and

(2) a second plasmid comprising BAV left or right end nucleotide sequences containing the E1 or E3 gene regions, respectively, having a heterologous nucleotide sequence encoding a foreign
30 gene or fragment thereof inserted for the deletion of a part or all of the E1 or E3 gene regions.

9. A live viable recombinant bovine adenovirus (BAV) comprising a deletion of part or all
35 of the E1 gene region, a deletion of part or all of the E3 gene region or deletion of both, and inserted into at least one deletion a heterologous nucleotide sequence coding for a polypeptide or an antigenic determinant produced by a disease causing organism.

-103-

10. A live viable recombinant bovine adenovirus (BAV) for producing an immune response in a mammalian host comprising:

- (1) a live bovine adenovirus (BAV)
5 modified to contain a heterologous nucleotide sequence coding for a polypeptide or an antigenic determinant corresponding to the desired immune response in association with or without
- (2) an effective promoter for said
10 nucleotide sequence to provide expression of said antigenic determinant in immunogenic non-pathogenic quantities.

11. A live recombinant bovine adenovirus
15 expression system comprising a deletion of all or part of the E1 gene region or all or part of the E3 gene region, or both deletions and inserted in at least one deletion a heterologous nucleotide sequence coding for a foreign gene or fragment thereof under control of an
20 expression promoter with or without one or more polyadenylation signal.

12. A recombinant mammalian cell line
comprising bovine adenovirus (BAV) E1 gene region,
25 said recombinant cell line thereby capable of allowing replication therein of a bovine adenovirus comprising an E1 deletion which may or may not be replaced by a heterologous or homologous nucleotide sequence encoding a foreign gene or fragment thereof.

30

13. The cell line of claim 12 which is a bovine cell line.

14. The recombinant mammalian cell line of
35 claim 12 wherein the heterologous or homologous nucleotide sequence encoding the foreign gene or fragment thereof is selected from the group consisting of a bovine adenovirus (BAV) E1 polypeptide,, a BAV

-104-

E1-associated polypeptide, a growth factor, a cellular receptor or other cellular polypeptide.

15. A recombinant mammalian cell line
- 5 comprising bovine adenovirus E1 genes, said recombinant cell line thereby capable of allowing DNA-mediated transfection to generate a recombinant bovine adenovirus (BAV) selected from the group consisting of:
- 10 (a) a recombinant BAV wherein part or all of the E1 gene region is replaced by a heterologous nucleotide sequence encoding a foreign gene or fragment thereof,
- (b) a recombinant BAV wherein part or all of
- 15 the E3 gene region is replaced by a heterologous nucleotide sequence encoding a foreign gene or fragment thereof,
- (c) a recombinant BAV wherein part or all of the E1 gene region and part or all of the E3 gene
- 20 region are deleted and inserted into at least one deletion a heterologous nucleotide sequence encoding a foreign gene or fragment thereof,
- (d) a recombinant BAV wherein part or all of the E1 gene region and/or part or all of the E3 gene
- 25 region are deleted and inserted into at least one deletion a heterologous nucleotide sequence encoding more than one foreign gene or fragment thereof to produce a recombinant fusion protein, and
- (e) a mutant BAV wherein part or all of the
- 30 E1 gene region and/or part or all of the E3 gene region are deleted.

16. A method of preparing a recombinant polypeptide or fragment thereof which comprises:
- 35 (1) infecting the mammalian cell line of claim 12, with a recombinant bovine adenovirus comprising a deletion of part or all of the E1 gene region and/or part or all of the E3 gene region and inserted into at least one deletion a heterologous

-105-

nucleotide sequence encoding the polypeptide or fragment thereof,

(2) replicating the recombinant virus in a recombinant cell line under conditions to provide for expression of the polypeptide, and

(3) recovering the recombinant polypeptide or antigenic fragment thereof.

17. A method of isolating a polypeptide which comprises:

(1) replicating a recombinant mammalian cell line of claim 12 under conditions to provide for expression of the polypeptide, and

(2) recovering the polypeptide or fragment thereof.

18. A method for eliciting an immune response in a mammalian host to protect against an infection comprising:

administering a vaccine composition comprising a live recombinant BAV of claim 1 wherein the foreign gene or fragment encodes an antigen with or without a pharmaceutically acceptable carrier.

25

19. A method for eliciting an immune response in a mammalian host to protect against an infection comprising:

administering a vaccine comprising a recombinant polypeptide or fragment thereof prepared by a method of claim 16 with or without a pharmaceutically acceptable carrier.

20. A vaccine for protecting a mammalian host against infection comprising a live recombinant adenovirus of claim 1 wherein the foreign gene or fragment encodes an antigen with or without a pharmaceutically acceptable carrier.

-106-

21. A vaccine for protecting a mammalian host against infection comprising a recombinant antigen prepared by a method of claim 16 with or without a pharmaceutically acceptable carrier.

5

22. A mutant bovine adenovirus (BAV) comprising a deletion of part or all of E1 and/or a deletion of part or all of E3.

10

23. A method for providing gene therapy to a mammal in need thereof to control a gene deficiency which comprises administering to said mammal a live recombinant bovine adenovirus containing a foreign nucleotide sequence encoding a non-defective form of said gene under conditions wherein the recombinant virus vector genome is incorporated into said mammalian genome or is maintained independently and extrachromosomally to provide expression of the required gene in the target organ or tissue.

20

25

30

35

1/51

```
10      20      30      40      50      60
CATCATCAAT AATCTACAGT AACTGATGG CAGCGGTCCA ACTGCCAATC ATTTTGGCCA

70      80      90      100     110     120
CGTCATTAT  GACGCAACGA CGGCGAGCGT GCGGTGCTGA CGTAACTGTG GGGCGGAGCG

130     140     150     160     170     180
CGTCGCGGAG GCGGCGGCGC TGGGCGGGGC TGAGGGCGGC GGGGGCGGCG CGCGGGGCGG

190     200     210     220     230     240
CGCGCGGGGC GGGGCGAGGG GCGGAGTTCC GCACCCGCTA CGTCATTTTC AGACATTTTT

250     260     270     280     290     300
TAGCAAAATT GCGCCTTTTG CAAGCATTTT TCTCACATTT CAGGTATTTA GAGGGCGGAT

310     320     330     340     350     360
TTTTGGTGTT CGTACTTCCG TGTCACATAG TTCACTGTCA ATCTTCATTA CGGCTTAGAC

370     380     390     400     410     420
AAATTTTCGG CGTCTTTTCC GGGTTTATGT CCCC GGTCAC CTTTAAGACT GTGTGAAACA

430     440     450     460     470     480
CACCTGCCCC TTGTTTACCC TTGGTCAGTT TTTTCGTCTC CTAGGGTGGG AACATCAAGA
```

FIG. 1A

2/51

```

490      500      510      520      530      540
ACAAATTGCGAGTAATTGTCACCTTTTCCGCGTTAGACTGCGTTT CACACGTAGA

550      560      570      580      590      600
CAGACTTTTCTCATTTTCTCACACTCCGTCGTCCGCTTCAGAGCTCTGC GTCTTCGCTG

610      620      630      640      650
CCACC ATG AAG TAC CTG GTC CTC GTT CTC AAC GAC GGC ATG AGT CGA ATT GAA
Met Lys Tyr Leu Val Leu Val Leu Asn Asp Gly Met Ser Arg Ile Glu

660      670      680      690      700
AAA GCT CTC CTG TGC AGC GAT GGT GAG GTG GAT TTA GAG TGT CAT GAG GTA
Lys Ala Leu Leu Cys Ser Asp Gly Glu Val Asp Leu Glu Cys His Glu Val

710      720      730      740      750
CTT CCC CCT TCT CCC GCG CCT GTC CCC GCT TCT GTG TCA CCC GTG AGG AGT
Leu Pro Pro Ser Pro Ala Pro Val Pro Ala Ser Val Ser Pro Val Arg Ser

760      770      780      790      800
CCT CCT CCT CTG TCT CCG GTG TTT CCT CCG TCT CCG CCA GCC CCG CTT GTG
Pro Pro Pro Leu Ser Pro Val Phe Pro Pro Ser Pro Pro Ala Pro Leu Val

810      820      830      840      850
AAT CCA GAG GCG AGT TCG CTG CTG CAG CAG TAT CCG AGA GAG CTG TTA GAG
Asn Pro Glu Ala Ser Ser Leu Leu Gln Gln Tyr Arg Arg Glu Leu Leu Glu

```

FIG. 1B

3/51

```

860      870      880      890      900
AGG AGC CTG CTC CGA ACG GCC GAA GGT CAG CAG CGT GCA GTG TGT CCA TGT
Arg Ser Leu Leu Arg Thr Ala Glu Gly Gln Gln Arg Ala Val Cys Pro Cys

910      920      930      940      950
GAG CCG TTG CCC GTG GAA GAG GAT GAG TGT CTG AAT GCC GTA AAT TTG CTG
Glu Arg Leu Pro Val Glu Glu Asp Glu Cys Leu Asn Ala Val Asn Leu Leu

960      970      980      990      1000      1010
TTT CCT GAT CCC TGG CTA AAT GCA GCT GAA AAT GGG GGT GAT ATT TTT AAG
Phe Pro Asp Pro Trp Leu Asn Ala Ala Glu Asn Gly Gly Asp Ile Phe Lys

1020      1030      1040      1050      1060
TCT CCG GCT ATG TCT CCA GAA CCG TGG ATA GAT TTG TCT AGC TAC GAT AGC
Ser Pro Ala Met Ser Pro Glu Pro Trp Ile Asp Leu Ser Ser Tyr Asp Ser

1070      1080      1090      1100      1110
GAT GTA GAA GAG GTG ACT AGT CAC TTT TTT CTG GAT TGC CCT GAA GAC CCC
Asp Val Glu Glu Val Thr Ser His Phe Phe Leu Asp Cys Pro Glu Asp Pro

1120      1130      1140      1150      1160
AGT CCG GAG TGT TCA TCT TGT GGG TTT CAT CAG GCT CAA AGC GGA ATT CCA
Ser Arg Glu Cys Ser Ser Cys Gly Phe His Gln Ala Gln Ser Gly Ile Pro

```

FIG. 1C

4/51

1170 1180 1190 1200 1210
 GGC ATT ATG TGC AGT TTG TGC TAC ATG CGC CAA ACC TAC CAT TGC ATC TAT
 Gly Ile Met Cys Ser Leu Cys Tyr Met Arg Gln Thr Tyr His Cys Ile Tyr

 1220 1230 1240 1250 1260 1270
 A[GTAAG TACATTCTGT AAAAGAACAT CTTCGTGATT TCTAGGTATT GTTTAGGGAT
 S

 1280 1290 1300 1310 1320 1330
 TAACTGGGTG GAGTGATCTT AATCCGGCAT AACCAATAC ATGTTTCAC AG]GT CCA GTT
 er Pro Val

 1340 1350 1360 1370 1380 1390
 TCT GAA GAG GAA ATG TGAGT CATGTTGACT TTGGCGCGC A AGAGGAAATG TGAGTCATGT
 Ser Glu Glu Glu Met End

 1400 1410 1420 1430 1440 1450
 TGACTTTGGC GCGCCCTACG GTGACTTTAA AGCAATTGA GGATCACTTT TTTGTTAGTC

 1460 1470 1480 1490 1500
 GCTATAAAGT AGTCACGGAG TCTTC ATG GAT CAC TTA AGC GTT CTT TTG GAT TTG
 Met Asp His Leu Ser Val Leu Leu Asp Leu

 1510 1520 1530 1540 1550
 AAG CTG CTT CGC TCT ATC GTA GCG GGG GCT TCA AAT CGC ACT GGA GTG TGG
 Lys Leu Leu Arg Ser Ile Val Ala Gly Ala Ser Asn Arg Thr Gly Val Trp

FIG. ID

5/51

1560 1570 1580 1590 1600
 AAG AGG CGG CTG TGG CTG GGA CGC CTG ACT CAA CTG GTC CAT GAT ACC TGC
 Lys Arg Arg Leu Trp Leu Gly Arg Leu Thr Gln Leu Val His Asp Thr Cys

1610 1620 1630 1640 1650
 GTA GAG AAC GAG AGC ATA TTT CTC AAT TCT CTG CCA GGG AAT GAA GCT TTT
 Val Glu Asn Glu Ser Ile Phe Leu Asn Ser Leu Pro Gly Asn Glu Ala Phe

1660 1670 1680 1690 1700
 TTA AGG TTG CTT CGG AGC GGC TAT TTT GAA GTG TTT GAC GTG TTT GTG GTG
 Leu Arg Leu Leu Arg Ser Gly Tyr Phe Glu Val Phe Asp Val Phe Val Val

1710 1720 1730 1740 1750 1760
 CCT GAG CTG CAT CTG GAC ACT CCG GGT CGA GTG GTC GCC GCT CTT GCT CTG
 Pro Glu Leu His Leu Asp Thr Pro Gly Arg Val Ala Ala Leu Ala Leu

1770 1780 1790 1800 1810
 CTG GTG TTC ATC CTC AAC GAT TTA GAC GCT AAT TCT GCT TCT TCA GGC TTT
 Leu Val Phe Ile Leu Asn Asp Leu Asp Ala Asn Ser Ala Ser Ser Gly Phe

1820 1830 1840 1850 1860
 GAT TCA GGT TTT CTC GTG GAC CGT CTC TGC GTG CCG CTA TGG CTG AAG GCC
 Asp Ser Gly Phe Leu Val Asp Arg Leu Cys Val Pro Leu Trp Leu Lys Ala
 Met Ala Glu Gly

FIG. 1E

6/51

1870	1880	1890	1900	1910
AGG GCG TTC AAG ATC ACC CAG AGC TCC AGG AGC ACT TCG CAG CCT TCC TCG				
Arg Ala Phe Lys Ile Thr Gln Ser Ser Arg Ser Thr Ser Gln Pro Ser Ser				
Gln Gly Val Gln Asp His Pro Glu Leu Gln Glu His Phe Ala Ala Phe Leu				
1920	1930	1940	1950	1960
TCG CCC GAC AAG ACG ACC CAG ACT ACC AGC CAG TA GAC GGG GAC AGC CCA				
Ser Pro Asp Lys Thr Thr Gln Thr Thr Ser Gln End				
Val Ala Arg Gln Asp Asp Pro Asp Tyr Gln Pro Val Asp Gly Asp Ser Pro				
1970	1980	1990	2000	2010
CCC CGG GCT AGC CTG GAG GAG GCT GAA CAG AGC AGC ACT CGT TTC GAG CAC				
Pro Arg Ala Ser Leu Glu Glu Ala Glu Gln Ser Ser Thr Arg Phe Glu His				
2020	2030	2040	2050	2060
ATC AGT TAC CGA GAC GTG GTG GAT GAC TTC AAT AGA TGC CAT GAT GTT TTT				
Ile Ser Tyr Arg Asp Val Val Asp Asp Phe Asn Arg Cys His Asp Val Phe				
2070	2080	2090	2100	2110
TAT GAG AGG TAC AGT TTT GAG GAC ATA AAG AGC TAC GAG GCT TTG CCT GAG				
Tyr Glu Arg Tyr Ser Phe Glu Asp Ile Lys Ser Tyr Glu Ala Leu Pro Glu				

FIG. IF

7/51

2120 2130 2140 2150 2160
 GAC AAT TTG GAG CAG CTC ATA GCT ATG CAT GCT AAA ATC AAG CTG CTG CCC
 Asp Asn Leu Glu Gln Leu Ile Ala Met His Ala Lys Ile Lys Leu Leu Pro

2170 2180 2190 2200 2210
 GGT CGG GAG TAT GAG TTG ACT CAA CCT TTG AAC ATA ACA TCT TGC GCC TAT
 Gly Arg Glu Tyr Glu Leu Thr Gln Pro Leu Asn Ile Thr Ser Cys Ala Tyr

2220 2230 2240 2250 2260
 GTG CTC GGA AAT GGG GCT ACT ATT AGG GTA ACA GGG GAA GCC TCC CCG GCT
 Val Leu Gly Asn Gly Ala Thr Ile Arg Val Thr Gly Glu Ala Ser Pro Ala

2270 2280 2290 2300 2310 2320
 ATT AGA GTG GGG GCC ATG GCC GTG GGT CCG TGT GTA ACA GGA ATG ACT GGG
 Ile Arg Val Gly Ala Met Ala Val Gly Pro Cys Val Thr Gly Met Thr Gly

2330 2340 2350 2360 2370
 GTG ACT TTT GTG AAT TGT AGG TTT GAG AGA GAG TCA ACA ATT AGG GGG TCC
 Val Thr Phe Val Asn Cys Arg Phe Glu Arg Glu Ser Thr Ile Arg Gly Ser

2380 2390 2400 2410 2420
 CTG ATA CGA GCT TCA ACT CAC GTG CTG TTT CAT GGC TGT TAT TTT ATG GGA
 Leu Ile Arg Ala Ser Thr His Val Leu Phe His Gly Cys Tyr Phe Met Gly

FIG. 1G

8/51

2430	2440	2450	2460	2470
ATT ATG GGC ACT TGT ATT GAG GTG GGG GCG GGA GCT TAC ATT CCG GGT TGT				
Ile Met Gly Thr Cys Ile Glu Val Gly Ala Gly Ala Tyr Ile Arg Gly Cys				
2480	2490	2500	2510	2520
GAG TTT GTG GGC TGT TAC CGG GGA ATC TGT TCT ACT TCT AAC AGA GAT ATT				
Glu Phe Val Gly Cys Tyr Arg Gly Ile Cys Ser Thr Ser Asn Arg Asp Ile				
2530	2540	2550	2560	2570
AAG GTG AGG CAG TGC AAC TTT GAC AAA TGC TTA CTG GGT ATT ACT TGT AAG				
Lys Val Arg Gln Cys Asn Phe Asp Lys Cys Leu Leu Gly Ile Thr Cys Lys				
2580	2590	2600	2610	2620
GGG GAC TAT CGT CTT TCG GGA AAT GTG TGT TCT GAG ACT TTC TGC TTT GCT				
Gly Asp Tyr Arg Leu Ser Gly Asn Val Cys Ser Glu Thr Phe Cys Phe Ala				
2630	2640	2650	2660	2670
CAT TTA GAG GGA GAG GGT TTG GTT AAA AAC AAC ACA GTC AAG TCC CCT AGT				
His Leu Glu Gly Glu Gly Leu Val Lys Asn Asn Thr Val Lys Ser Pro Ser				
2680	2690	2700	2710	2720
CGC TGG ACC AGC GAG TCT GGC TTT TCC ATG ATA ACT TGT GCA GAC GGC AGG				
Arg Trp Thr Ser Glu Ser Gly Phe Ser Met Ile Thr Cys Ala Asp Gly Arg				

FIG. 1H

9/51

2730 2740 2750 2760 2770
 GTT ACG CCT TTG GGT TCC CTC CAC ATT GTG GGC AAC CGT TGT AGG CGT TGG
 Val Thr Pro Leu Gly Ser Leu His Ile Val Gly Asn Arg Cys Arg Arg Trp

2780 2790 2800 2810 2820 2830
 CCA ACC ATG CAG GGG AAT GTG TTT ATC ATG TCT AAA CTG TAT CTG GGC AAC
 Pro Thr Met Gln Gly Asn Val Phe Ile Met Ser Lys Leu Tyr Leu Gly Asn

2840 2850 2860 2870 2880
 AGA ATA GGG ACT GTA GCC CTG CCC CAG TGT GCT TTC TAC AAG TCC AGC ATT
 Arg Ile Gly Thr Val Ala Leu Pro Gln Cys Ala Phe Tyr Lys Ser Ser Ile

2890 2900 2910 2920 2930
 TGT TTG GAG GAG AGG GCG ACA AAC AAG CTG GTC TTG GCT TGT GCT TTT GAG
 Cys Leu Glu Glu Arg Ala Thr Asn Lys Leu Val Leu Ala Cys Ala Phe Glu

2940 2950 2960 2970 2980
 AAT AAT GTA CTG GTG TAC AAA GTG CTG AGA CGG GAG AGT CCC TCA ACC GTG
 Asn Asn Val Leu Val Tyr Lys Val Leu Arg Arg Glu Ser Pro Ser Thr Val

2990 3000 3010 3020 3030
 AAA ATG TGT GTT TGT GGG ACT TCT CAT TAT GCA AAG CCT TTG ACA CTG GCA
 Lys Met Cys Val Cys Gly Thr Ser His Tyr Ala Lys Pro Leu Thr Leu Ala

FIG. 11

10/51

```

3040      3050      3060      3070      3080
ATT ATT TCT TCA GAT ATT CGG GCT AAT CGA TAC ATG TAC ACT GTG GAC TCA
Ile Ile Ser Ser Asp Ile Arg Ala Asn Arg Tyr Met Tyr Thr Val Asp Ser

3090      3100      3110      3120      3130      3140
ACA GAG TTC ACT TCT GAC GAG GAT T AAAAGTGGC GGGGCCAAGA GGGGTATAAA
Thr Glu Phe Thr Ser Asp Glu Asp End

3150      3160      3170      3180      3190      3200
TAGGTGGGA GGTGAGGG AGCCGTAGTT TCTGTTTTC CCAGACTGGG GGGGACAAC ATG
Met

3210      3220      3230      3240      3250
GCC GAG GAA GGG CGC ATT TAT GTG CCT TAT GTA ACT GCC CGC CTG CCC AAG
Ala Glu Glu Gly Arg Ile Tyr Val Pro Tyr Val Thr Ala Arg Leu Pro Lys

3260      3270      3280      3290      3300
TGG TCG GGT TCG GTG CAG GAT AAG ACG GGC TCG AAC ATG TTG GGG GGT GTG
Trp Ser Gly Ser Val Gln Asp Lys Thr Gly Ser Asn Met Leu Gly Gly Val

3310      3320      3330      3340      3350
GTA CTC CCT CCT AAT TCA CAG GCG CAC CGG ACG GAG ACC GTG GGC ACT GAG
Val Leu Pro Pro Asn Ser Gln Ala His Arg Thr Glu Thr Val Gly Thr Glu

```

FIG. 1J

11/51

```

3360      3370      3380      3390      3400
GCC ACC AGA GAC AAC CTG CAC GCC GAG GGA GCG CGT CGT CCT GAG GAT CAG
Ala Thr Arg Asp Asn Leu His Ala Glu Gly Ala Arg Arg Pro Glu Asp Gln

3410      3420      3430      3440      3450
ACG CCC TAC ATG ATC TTG GTG GAG GAC TCT CTG GGA GGT TTG AAG AGG CGA
Thr Pro Tyr Met Ile Leu Val Glu Asp Ser Leu Gly Gly Leu Lys Arg Arg

3460      3470      3480      3490      3500
ATG GAC TTG CTG GAA GAA TCT AAT CAG CAG CTG CTG GCA ACT CTC AAC CGT
Met Asp Leu Leu Glu Glu Ser Asn Gln Gln Leu Leu Ala Thr Leu Asn Arg

3510      3520      3530      3540      3550
CTC CGT ACA GGA CTC GCT GCC TAT GTG CAG GCT AAC CTT GTG GGC GGC CAA
Leu Arg Thr Gly Leu Ala Ala Tyr Val Gln Ala Asn Leu Val Gly Gly Gln

3560      3570      3580      3590      3600      3610
GTT AAC CCC TTT GTT TAAATA AAAATACACT CATACAGTTT ATTATGCTGT
Val Asn Pro Phe Val End

3620      3630      3640      3650      3660      3670
CAATAAAATT CTTTATTTT CCTGTGATAA TACCGTGTC AGCGTGCTCT GTCAATAAGG

3680      3690      3700      3710      3720      3730
GTCCTATGCA TCCTGAGAAG GGCCTCATAT ACCCATGGCA TGAATATTA GATACATGGG

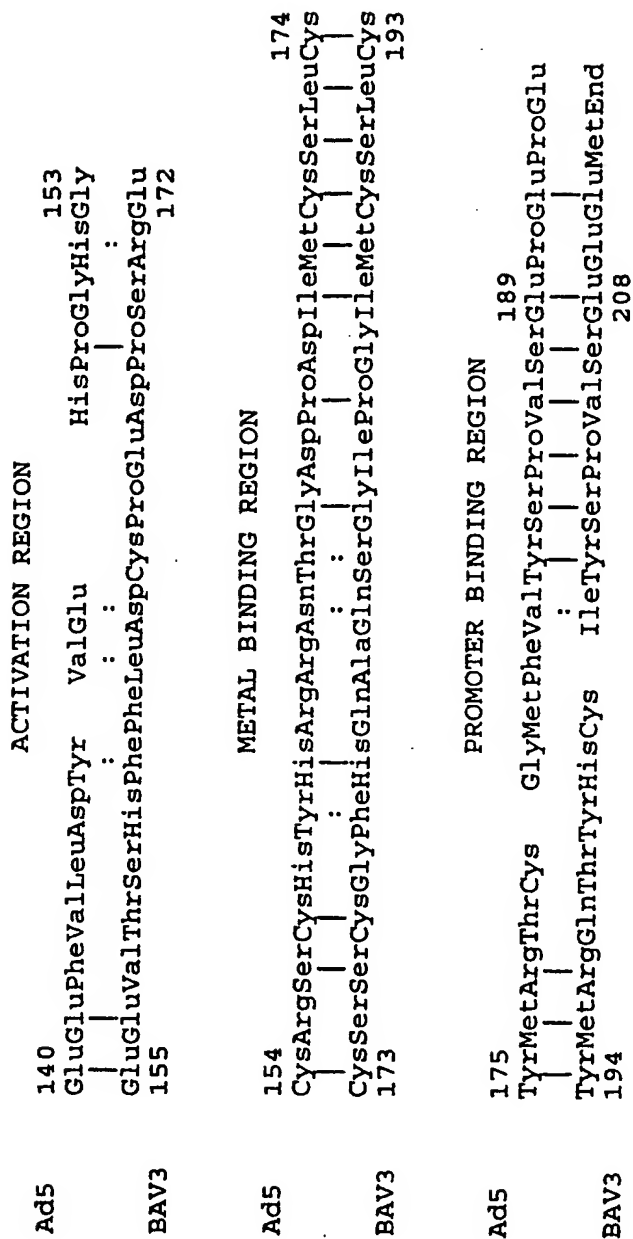
```

FIG. 1K

12/51

3740 3750 3760 3770 3780 3790
CATAAGGCC TCAGAGGGT TGAGGTAGAG CCACTGCAGA CTTTCGTGG GAGGTAAGGT
3800 3810 3820 3830 3840 3850
GTTGTAAATA ATCCAGTCAT ACTGACTGTG CTGGGCGTGG AAGGAAAAGA TGTCTTTTAG
3860 3870 3880 3890 3900 3910
AAGAAGGGTG ATTGGCAAAG GGAGGCTCTT AGTGTAGGTA TTGATAAATC TGTTCAAGTTG
3920 3930 3940 3950 3960 3970
GGAGGGATGC ATTCGGGGGC TAATAAGGTG GAGTTTAGCC TGAATCTTAA GGTTGGCAAT
3980 3990 4000 4010 4020 4030
GTTGCCCCCT AGGTCTTTGC GAGGATTTCAT GTTGTGCAGT ACCACAAAAA CAGAGTAGCC
4040 4050 4060
TGTCATTG GGAATTAT CATGAAGCT T

FIG. 1L



14/51

Rb BINDING SEQUENCE

Ad5	120	IleAspLeuThrCysHisGluAlaGlyPheProProSer	132
		:	
BAV3	26	ValAspLeuGluCysHisGluVal LeuProProSer	37

FIG. 2B

Ad5	82	LeuAspPheSerThrProGlyArgAlaAlaAlaAlaValAlaPheLeuSerPheIle	100
BAV3	83	LeuAsp ThrProGlyArgValValAlaAlaLeuAlaLeuLeuValPheIle	99

FIG. 3A

Ad5	20	GlnSerSerAsnSerThrSer	26
BAV3	136	GlnSerSerArgSerThrSer	142

FIG. 3B

15/51

```

Ad5 150  GlnLysTyrSerIleGluGlnLeuThrThrTyrTrpLeuGlnProGlyAspAspPheGlu
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
BAV3 74  GluArgTyrLysPheGluAspIleLysSerTyrGluAlaLeuProGluAspAsnLeuGlu
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
170      GluAlaIleArgValTyrAlaLysValAlaLeuArgProAspCysLysTyrLysIleSer
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
94      GlnLeuIleAlaMetHisAlaLysIleLysLeuLeuProGlyArgGluTyrGluLeuThr
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
190      LysLeuValAsnIleArgAsnCysCysTyrIleSerGlyAsnGlyAlaGluValGluIle
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
114      GlnProLeuAsnIleThrSerCysAlaTyrValLeuGlyAsnGlyAlaThrIleArgVal
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
210      AspThrGluAspArgValAlaPheArgCysSerMetIleAsnMetTrpProGlyValLeu
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
134      ThrGlyGluAlaSerProAlaIleArgValGlyAlaMetAlaValGlyProCysValThr
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
230      GlyMetAspGlyValValIleMetAsnValArgPheThr GlyProAsnPheSerGly
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
154      GlyMetThrGlyValThrPheValAsnCysArgPheGluArgGluSerThrIleArgGly
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
249      ThrValPheLeuAlaAsnThrAsnLeuIleLeuHisGlyValSerPheTyr GlyPhe
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
174      SerLeuIleArgAlaSerThrHisValLeuPheHisGlyCys TyrPheMetGlyIle
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
268      AsnAsnThrCysValGluAlaTrpThrAspValArgValArgGlyCysAlaPheTyrCys
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
193      MetGlyThrCysIleGluValGlyAlaGlyAlaTyrIleArgGlyCysGluPheValGly
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

FIG. 4A

16/51

```

288 CysTrpLysGlyValValCysArgProLysSerArgAla SerIleLysLysCysLeu
      : : :
213 CysTyrArgGlyIle CysSerThrSerAsnArgAspIleLysValArgGlnCysAsn
      : : :
307 PheGluArgCysThrLeuGlyIleLeuSerGluGlyAsnSerArgValArgHisAsnVal
      : : :
232 PheAspLysCysLeuLeuGlyIleThrCysLysGlyAspTyrArgLeuSerGlyAsnVal
      : : :
327 AlaSerAspCysGlyCysPheMetLeuValLysSerValAlaValIleLysHisAsnMet
      : : :
252 CysSerGluThrPheCysPheAlaHisLeuGluGlyGluGlyLeuValLysAsnAsnThr
      : : :
347 Val CysGlyAsn CysGluAspArgAlaSerGlnMetLeuThrCysSerAsp
      : : :
272 ValLysSerProSerArgTrpThrSerGluSerGlyPheSerMetIleThrCysAlaAsp
      : : :
364 GlyAsnCysHisLeuLeuLysThrIleHisVal AlaSerHisSerArgLysAlaTrp
      : : :
292 GlyArgValThrProLeuGlySerLeuHisIleValGlyAsnArgCysArgArg Trp
      : : :
383 ProValPheGluHisAsnIleLeuThrArgCysSerLeuHisLeuGlyAsnArgArgGly
      : : :
311 ProThrMetGlnGlyAsnValPheIleMetSerLysLeuTyrLeuGlyAsnArgIleGly
      : : :
403 ValPheLeuProTyrGlnCysAsnLeuSerHisThrLysIleLeuLeuGluProGlu
      : : :
331 ThrValAlaLeuPro GlnCysAlaPheTyrLysSerSerIleCysLeuGluGluArg
      : : :

```

FIG. 4B

17/51

```

422 SerMetSerLysValAsnLeuAsnGlyValPheAspMetThrMetLysIleTrpLysVal
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
350 AlaThrAsnLysLeuValLeuAlaCysAlaPheGluAsnAsnValLeuValTyrLysVal
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
442 LeuArgTyrAspGluThrArgThrArgCysArgProCysGluCysGlyGlyLysHisIle
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
370 LeuArgArgGluSerProSerThr  VallysMetCysValCysGlyThrSerHisTyr
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
462 ArgAsnGlnProValMetLeuAspVal  ThrGluGluLeuArgProAspHisLeuVal
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
389 AlaLysProLeuThrLeuAlaIleIleSerSerAspIleArgAlaAsnArgTyrMet
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
481 LeuAlaCysThrArgAlaGluPheGlySerSerSerAspGluAspThrAspEnd
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
408 TyrThrValAspSerThrGluPhe  ThrSerAspGluAspEnd
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

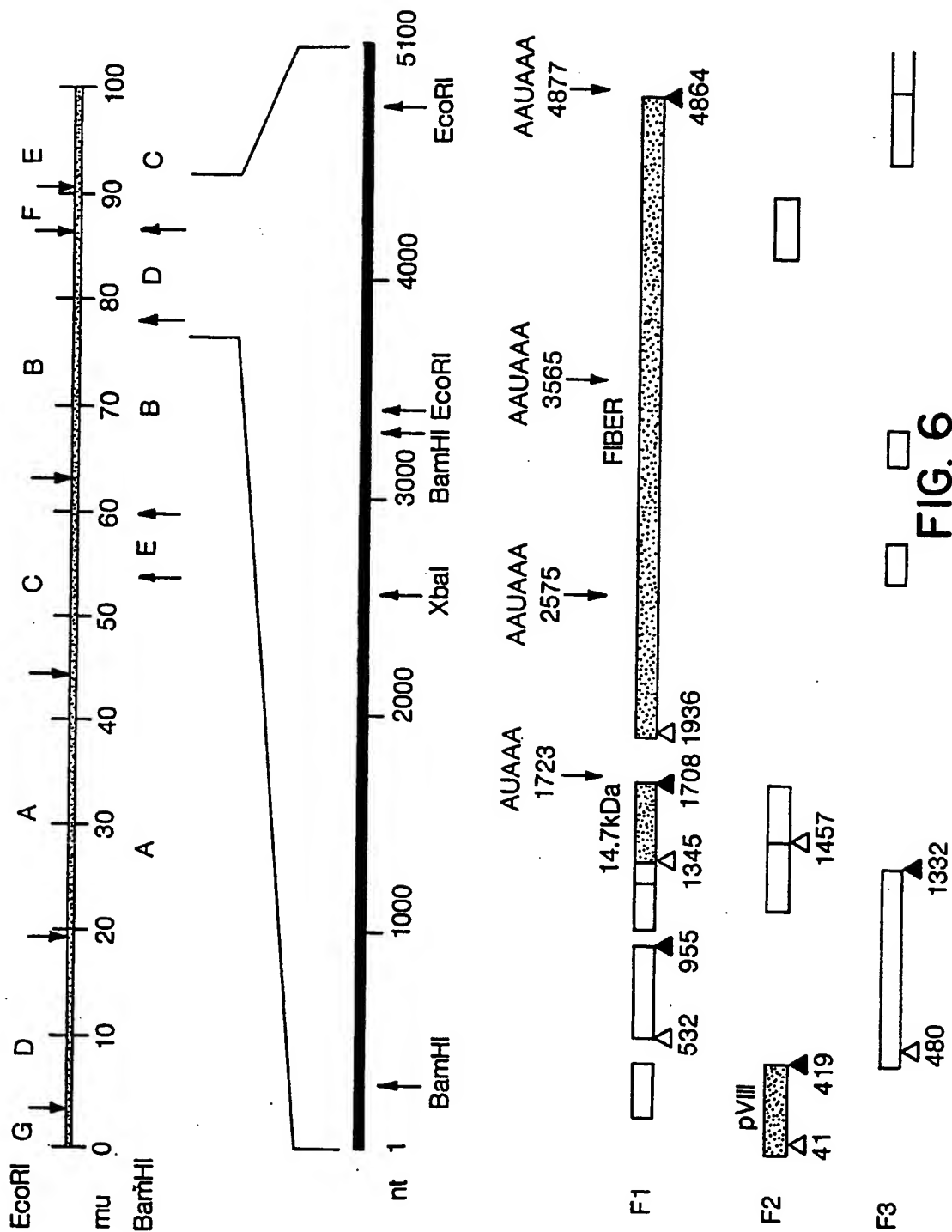
FIG. 4C

18/51

Ad5	1	MetSerThrAsnSerPheAspGlySerIleValSerSerTyrLeuThrThrArgMetPro
		:
BAV3	1	MetAla Glu GluGlyArgIleTyrValProTyrValThrAlaAlaArgLeuPro
		:
	21	ProTrpAlaGlyValArgGlnAsnValMetGlySerSerIleAspGlyArgProValLeu
		:
	18	LysTrpSerGlySerValGlnAspLysThrGlySerAsnMetLeuClyGlyValValLeu
		:
	41	ProAlaAsnSerThrThrLeuThrTyrGluThrValSerGlyThrProLeuGluThrAla
	38	ProProAsnSerGlnAlaHisArgThrGluThrVal GlyThrGlu AlaThr
		:
	61	AlaSerAlaAlaAlaSerAlaAlaAlaAlaThrAlaArgGlyIleValThrAspPheAla
	55	ArgAspAsnLeuHisAlaGluGlyAlaAlaArg ArgProGluAspGlnThr Pro
	81	PheLeuSerProLeuAlaSerSerAlaAlaSerArgSerSerAlaArgAspAspLysLeu
		:
	72	TyrMetIle LeuValGluAspSerLeuGlyGlyLeuLysArgArgMetAspLeuLeu
	101	ThrAlaLeuLeuAlaGlnLeu AspSerLeuThrArgGluLeuAsnValValSerGln
	91	GluGluSerAsnGlnGlnLeuLeuAlaThrLeuAsnArg LeuArgThr Gly
	120	GlnLeuLeuAspLeuArgGlnGlnValSerAlaLeuLysAlaSerSerProProAsnAla
	108	LeuAlaAlaTyr ValGln AlaAsnLeuValGlyGlyGlnValAsnProphe
	140	ValEnd
	125	ValEnd

FIG. 5

19/51



20/51

10 20 30 40 50
 C CTC ATC AAA CAA CCC GTG GTG GGC ACC ACC CAC GTG GAA ATG CCT CGC AAC
 ORF 1 Leu Ile Lys Gln Pro Val Val Gly Thr Thr His Val Glu Met Pro Arg Asn

 60 70 80 90 100
 GAA GTC CTA GAA CAA CAT CTG ACC TCA CAT GGC GCT CAA ATC GCG GGC GGA
 Glu Val Leu Glu Gln His Leu Thr Ser His Gly Ala Gln Ile Ala Gly Gly

 110 120 130 140 150
 GGC GCT GCG GGC GAT TAC TTT AAA AGC CCC ACT TCA GCT CGA ACC CTT ATC
 Gly Ala Ala Gly Asp Tyr Phe Lys Ser Pro Thr Ser Ala Arg Thr Leu Ile

 160 170 180 190 200
 CCG CTC ACC GCC TCC TGC TTA AGA CCA GAT GGA GTC TTT CAA CTA GGA GGA
 Pro Leu Thr Ala Ser Cys Leu Arg Pro Asp Gly Val Phe Gln Leu Gly Gly

 210 220 230 240 250
 GGC TCG CGT TCA TCT TTC AAC CCC CTG CAA ACA GAT TTT GCC TTC CAC GCC
 Gly Ser Arg Ser Ser Phe Asn Pro Leu Gln Thr Asp Phe Ala Phe His Ala

 260 270 280 290 300
 CTG CCC TCC AGA CCG CGC CAC GGG GGC ATA GGA TCC AGG CAG TTT GTA GAG
 Leu Pro Ser Arg Pro Arg His Gly Gly Ile Gly Ser Arg Gln Phe Val Glu

FIG. 7A

21/51

310 320 330 340 350
 GAA TTT GTG CCC GCC GTC TAC CTC AAC CCC TAC TCG GGA CCG CCG GAC TCT
 Glu Phe Val Pro Ala Val Tyr Leu Asn Pro Tyr Ser Gly Pro Pro Asp Ser

360 370 380 390 400
 TAT CCG GAC CAG TTT ATA CGC CAC TAC AAC GTG TAC AGC AAC TCT GTG AGC
 Tyr Pro Asp Gln Phe Ile Arg His Tyr Asn Val Tyr Ser Asn Ser Val Ser

410 420 430 440 450 460
 GGT TAT AGC T GAG ATT GTA AGA CTC TCC TAT CTG TCT CTG TGC TGC TTT TCC
 Gly Tyr Ser

Val Ile Ala Glu Ile Val Arg Leu Ser Tyr Leu Ser Leu Cys Cys Phe Ser

470 480 490 500 510
 GCT TCA AGC CCC ACA AGC ATG AAG GGG TTT CTG CTC ATC TTC AGC CTG CTT
 Ala Ser Ser Pro Thr Ser Met Lys Gly Phe Leu Leu Ile Phe Ser Leu Leu

520 530 540 550 560
 GTG CAT TGT CCC CTA ATT CAT GTT GGG ACC ATT AGC TTC TAT GCT GCA AGG
 Val His Cys Pro Leu Ile His Val Gly Thr Ile Ser Phe Tyr Ala Ala Arg

ORF 3 Phe Met Leu Gly Pro Leu Ala Ser Met Leu Gln Gly

ORF 2 Ala

FIG. 7B

22/51

570 580 590 600 610
 CCC GGG TCT GAG CCT AAC GCG ACT TAT GTT TGT GAC TAT GGA AGC GAG TCA
 Pro Gly Leu Ser Leu Thr Arg Leu Met Phe Val Thr Met Glu Ala Ser Gln
 Pro Gly Ser Glu Pro Asn Ala Thr Tyr Val Cys Asp Tyr Gly Ser Glu Ser

620 630 640 650 660
 GAT TAC AAC CCC ACC ACG GTT CTG TGG TTG GCT CGA GAG ACC GAT GGC TCC
 Ile Thr Thr Pro Pro Arg Phe Cys Gly Trp Leu Glu Arg Pro Met Ala Pro
 Asp Tyr Asn Pro Thr Thr Val Leu Trp Leu Ala Arg Glu Thr Asp Gly Ser

670 680 690 700 710
 TGG ATC TCT GTT CTT TTC CGT CAC AAC GGC TCC TCA ACT GCA GCC CCC GGG
 Gly Ser Leu Phe Phe Ser Val Thr Thr Ala Pro Gln Leu Gln Pro Pro Gly
 Trp Ile Ser Val Leu Phe Arg His Asn Gly Ser Ser Thr Ala Ala Pro Gly

720 730 740 750 760
 GTC GTC GCG CAC TTT ACT GAC CAC AAC AGC AGC ATT GTG GTG CCC CAG TAT
 Ser Ser Arg Thr Leu Leu Thr Thr Thr Ala Ala Leu Trp Cys Pro Ser Ile
 Val Val Ala His Phe Thr Asp His Asn Ser Ser Ile Val Val Pro Gln Tyr

770 780 790 800 810
 TAC CTC CTC AAC AAC TCA CTC TCT AAG CTC TGC TGC TCA TAC CGG CAC AAC
 Thr Ser Ser Thr Thr His Ser Leu Ser Ser Ala Ala His Thr Gly Thr Thr
 Tyr Leu Leu Asn Asn Ser Leu Ser Lys Leu Cys Cys Ser Tyr Arg His Asn

FIG. 7C

23/51

820 830 840 850 860
 GAG CGT TCT CAG TTT ACC TGC AAA CAA GCT GAC GTC CCT ACC TGT CAC GAG
 Ser Val Leu Ser Leu Pro Ala Asn Lys Leu Thr Ser Leu Pro Val Thr Ser
 Glu Arg Ser Gln Phe Thr Cys Lys Gln Ala Asp Val Pro Thr Cys His Glu

870 880 890 900 910 920
 CCC GGC AAG CCG CTC ACC CTC CGC GTC TCC CCC GCG CTG GGA ACT GCC CAC
 Pro Ala Ser Arg Ser Pro Ser Ala Ser Pro Pro Arg Trp Glu Leu Pro Thr
 Pro Gly Lys Pro Leu Thr Leu Arg Val Ser Pro Ala Leu Gly Thr Ala His

930 940 950 960 970
 CAA GCA GTC ACT TGG TTT TTT CAA AAT GTA CCC ATA GCT ACT GTT TAC CGA
 Lys Gln Ser Leu Gly Phe Phe Lys Met Tyr Pro
 Gln Ala Val Thr Trp Phe Phe Gln Asn Val Pro Ile Ala Thr Val Tyr Arg

980 990 1000 1010 1020
 CCT TGG GGC AAT GTA ACT TGG TTT TGT CCT CCC TTC ATG TGT ACC TTT AAT
 Pro Trp Gly Asn Val Thr Trp Phe Cys Pro Pro Phe Met Cys Thr Phe Asn

1030 1040 1050 1060 1070
 GTC AGC CTG AAC TCC CTA CTT ATT TAC AAC TTT TCT GAC AAA ACC GGG GGG
Val Ser Leu Asn Ser Leu Leu Ile Tyr Asn Phe Ser Asp Lys Thr Gly Gly

FIG. 7D

24/51

1080	1090	1100	1110	1120
CAA TAC ACA GCT CTC ATG CAC TCC GGA CCT GCT TCC CTC TTT CAG CTC TTT				
Gln Tyr Thr Ala Leu Met His Ser Gly Pro Ala Ser Leu Phe Gln Leu Phe				
1130	1140	1150	1160	1170
AAG CCA ACG ACT TGT GTC ACC AAG GTG GAG GAC CCG CCG TAT GCC AAC GAC				
Lys Pro Thr Thr Cys Val Thr Lys Val Glu Asp Pro Pro Tyr Ala Asn Asp				
1180	1190	1200	1210	1220
CCG GCC TCG CCT GTG TGG CGC CCA CTG CTT TTT GCC TTC GTC CTC TGC ACC				
Pro Ala Ser Pro Val Trp Arg Pro Leu Leu Phe Ala Phe Val Leu Cys Thr				
1230	1240	1250	1260	1270
GGC TGC GCG GTG TTG TTA ACC GCC TTC GGT CCA TCG ATT CTA TCC GGT ACC				
Gly Cys Ala Val Leu Leu Thr Ala Phe Gly Pro Ser Ile Leu Ser Gly Thr				
ORF 4 Pro Pro Ser Val His Arg Phe Tyr Pro Val Pro				
1280	1290	1300	1310	1320
CGA AAG CTT ATC TCA GCC CGC TTT TGG AGT CCC GAG CCC TAT ACC ACC CTC				
Glu Ser Leu Ser Gln Pro Ala Phe Gly Val Pro Ser Pro Ile Pro Pro Ser				
Arg Lys Leu Ile Ser Ala Arg Phe Trp Ser Pro Glu Pro Tyr Thr Thr Leu				

FIG. 7E

25/51

1330 1340 1350 1360 1370 1380
CAC T AAC AGT CCC CCC ATG GAG CCA GAC GGA GTT CAT GCC GAG CAG CAG TTT
Thr Asn Ser Pro Pro Met Glu Pro Asp Gly Val His Ala Glu Gln Gln Phe
His

1390 1400 1410 1420 1430
ATC CTC AAT CAG ATT TCC TGC GCC AAC ACT GCC CTC CAG CGT CAA AGG GAG
Ile Leu Asn Gln Ile Ser Cys Ala Asn Thr Ala Leu Gln Arg Gln Arg Glu

1440 1450 1460 1470 1480
GAA CTA GCT TCC CTT GTC ATG TTG CAT GCC TGT AAG CGT GGC CTC TTT TGT
Glu Leu Ala Ser Leu Val Met Leu His Ala Cys Lys Arg Gly Leu Phe Cys
ORF 5 Leu Pro Leu Ser Cys Cys Met Pro Val Ser Val Ala Ser Phe Val

1490 1500 1510 1520 1530
CCA GTC AAA ACT TAC AAG CTC AGC CTC AAC GCC TCG GCC AGC GAG CAC AGC
Pro Val Lys Thr Tyr Lys Leu Ser Leu Asn Ala Ser Ala Ser Glu His Ser
Gln Ser Lys Leu Thr Ser Ser Ala Ser Thr Pro Arg Pro Ala Ser Thr Ala

1540 1550 1560 1570 1580
CTG CAC TTT GAA AAA AGT CCC TCC CGA TTC ACC CTG GTC AAC ACT CAC GCC
Leu His Phe Glu Lys Ser Pro Ser Arg Phe Thr Leu Val Asn Thr His Ala
Cys Thr Leu Lys Lys Val Pro Pro Asp Ser Pro Trp Ser Thr Leu Thr Pro

FIG. 7F

26/51

```

1590      1600      1610      1620      1630
GGA GCT TCT GTG CGA GTG GCC CTA CAC CAC GGT TCC GGC AGC ATC
Gly Ala Ser Val Arg Val Ala Leu His His Gln Gly Ala Ser Gly Ser Ile
Glu Leu Leu Cys Glu Trp Pro Tyr Thr Thr Arg Glu Leu Pro Ala Ala Ser

1640      1650      1660      1670      1680
CGC TGT TCC TGT TCC CAC GCC GAG TGC CTC CCC GTC CTC AAG ACC CTC
Arg Cys Ser Cys Ser His Ala Glu Cys Leu Pro Val Leu Lys Thr Leu
Ala Val Pro Val Pro Thr Pro Ser Ala Ser Pro Ser Ser Arg Pro Ser

1690      1700      1710      1720      1730      1740
TGT GCC TTT AAC TTT TTA GAT TAG CTGAAAGCAA ATATAAATG GTGTGCTTAC
Cys Ala Phe Asn Phe Leu Asp
Val Pro Leu Thr Phe

1750      1760      1770      1780      1790
CGTAATCTG TTTTGACTTG TGTGCTTGA TTT CTC CCC CTG CGC CGT AAT CCA GTG

1800      1810      1820      1830      1840
CCC CTC TTC AAA ACT CTC GTA CCC TAT GCG ATT CGC ATA GGC ATA TTT TCT

1850      1860      1870      1880      1890
AAA AGC TCT GAA GTC AAC ATC ACT CTC AAA CAC TTC TCC GTT GTA GGT TAC

```

FIG. 76

27/51

1900 1910 1920 1930 1940 1950
 TTT CAT CTA ATA AAG TCA TCC ACC GGT T AAC ATC ATG AAG AGA AGT GTG
 ORF 6 Ser His Pro Pro Val Asn Ile Met Lys Arg Ser Val

1960 1970 1980 1990 2000
 CCC CAG GAC TTT AAT CTT GTG TAT CCG TAC AAG GCT AAG AGG CCC AAC ATC
 Pro Gln Asp Phe Asn Leu Val Tyr Pro Tyr Lys Ala Lys Arg Pro Asn Ile

2010 2020 2030 2040 2050
 ATG CCG CCC TTT TTT GAC CGC AAT GGC TTT GTT GAA AAC CAA GAA GCC ACG
 Met Pro Pro Phe Phe Asp Arg Asn Gly Phe Val Glu Asn Gln Glu Ala Thr

2060 2070 2080 2090 2100
 CTA GCC ATG CTT GTG GAA AAG CCG CTC ACG TTC GAC AAG GAA GGT GCG CTG
 Leu Ala Met Leu Val Glu Lys Pro Leu Thr Phe Asp Lys Glu Gly Ala Leu

2110 2120 2130 2140 2150
 ACC CTG GGC GTC GGA CGC GGC ATC CGC ATT AAC CCC GCG GGG CTT CTG GAG
 Thr Leu Gly Val Gly Arg Gly Ile Arg Ile Asn Pro Ala Gly Leu Leu Glu

2160 2170 2180 2190 2200
 ACA AAC GAC CTC GCG TCC GCT GTC TTC CCA CCG CTG GCC TCC GAT GAG GCC
 Thr Asn Asp Leu Ala Ser Ala Val Phe Pro Pro Leu Ala Ser Asp Glu Ala

FIG. 7H

28/51

2210 2220 2230 2240 2250
 GGC AAC GTC ACG CTC AAC ATG TCT GAC GGG CTA TAT ACT AAG GAC AAC AAG
 Gly Asn Val Thr Leu Asn Met Ser Asp Gly Leu Tyr Thr Lys Asp Asn Lys

2260 2270 2280 2290 2300
 CTA GCT GTC AAA GTA GGT CCC GGG CTG TCC CTC GAC TCC AAT AAT GCT CTC
 Leu Ala Val Lys Val Gly Pro Gly Leu Ser Leu Asp Ser Asn Asn Ala Leu

2310 2320 2330 2340 2350
 CAG GTC CAC ACA GGC GAC GGG CTC ACG GTA ACC GAT GAC AAG GTG TCT CTA
 Gln Val His Thr Gly Asp Gly Leu Thr Val Thr Asp Asp Lys Val Ser Leu

2360 2370 2380 2390 2400
 AAT ACC CAA GCT CCC CTC TCG ACC ACC AGC GCG GGC CTC TCC CTA CTT CTG
 Asn Thr Gln Ala Pro Leu Ser Thr Thr Ser Ala Gly Leu Ser Leu Leu Leu

2410 2420 2430 2440 2450 2460
 GGT CCC AGC CTC CAC TTA GGT GAG GAG GAA CGA CTA ACA GTA AAC ACC GGA
 Gly Pro Ser Leu His Leu Gly Glu Glu Arg Leu Thr Val Asn Thr Gly

2470 2480 2490 2500 2510
 GCG GGC CTC CAA ATT AGC AAT AAC GCT CTG GCC GTA AAA GTA GGT TCA GGT
 Ala Gly Leu Gln Ile Ser Asn Asn Ala Leu Ala Val Lys Val Gly Ser Gly

FIG. 7I

29/51

2520	2530	2540	2550	2560
ATC ACC GTA GAT GCT CAA AAC CAG CTC GCT GCA TCC CTG GGG GAC GGT CTA				
Ile Thr Val Asp Ala Gln Asn Gln Leu Ala Ala Ser Leu Gly Asp Gly Leu				
2570	2580	2590	2600	2610
GAA AGC AGA GAT <u>AAT AAA</u> ACT GTC GTT AAG GCT GGG CCC GGA CTT ACA ATA				
Glu Ser Arg Asp <u>Asn Lys Thr</u> Val Val Lys Ala Gly Pro Gly Leu Thr Ile				
2620	2630	2640	2650	2660
ACT AAT CAA GCT CTT ACT GTT GCT ACC GGG AAC GGC CTT CAG GTC AAC CCG				
Thr Asn Gln Ala Leu Thr Val Ala Thr Gly Asn Gly Leu Gln Val Asn Pro				
2670	2680	2690	2700	2710
GAA GGG CAA CTG CAG CTA AAC ATT ACT GCC GGT CAG GGC CTC AAC TTT GCA				
Glu Gly Gln Leu Gln Leu <u>Asn Ile Thr</u> Ala Gly Gln Gly Leu Asn Phe Ala				
2720	2730	2740	2750	2760
AAC AAC AGC CTC GCC GTG GAG CTG GGC TCG GGC CTG CAT TTT CCC CCT GGC				
<u>Asn Asn Ser</u> Leu Ala Val Glu Leu Gly Ser Gly Leu His Phe Pro Pro Gly				
2770	2780	2790	2800	2810
CAA AAC CAA GTA AGC CTT TAT CCC GGA GAT GGA ATA GAC ATC CGA GAT AAT				
Gln Asn Gln Val Ser Leu Tyr Pro Gly Asp Gly Ile Asp Ile Arg Asp Asn				

FIG. 7J

30/51

2820 2830 2840 2850 2860
 AGG GTG ACT GTG CCC GCT GGG CCA GGC CTG AGA ATG CTC AAC CAC CAA CTT
 Arg Val Thr Val Pro Ala Gly Pro Gly Leu Arg Met Leu Asn His Gln Leu

2870 2880 2890 2900 2910
 GCC GTA GCT TCC GGA GAC GGT TTA GAA GTC CAC AGC GAC ACC CTC CGG TTA
 Ala Val Ala Ser Gly Asp Gly Leu Glu Val His Ser Asp Thr Leu Arg Leu

2920 2930 2940 2950 2960 2970
 AAG CTC TCC CAC GGC CTG ACA TTT GAA AAT GGC GCC GTA CGA GCA AAA CTA
 Lys Leu Ser His Gly Leu Thr Phe Glu Asn Gly Ala Val Arg Ala Lys Leu

2980 2990 3000 3010 3020
 GGA CCA GGA CTT GGC ACA GAC GAC TCT GGT CGG TCC GTG GTT CGC ACA GGT
 Gly Pro Gly Leu Gly Thr Asp Asp Ser Gly Arg Ser Val Val Arg Thr Gly

3030 3040 3050 3060 3070
 CGA GGA CTT AGA GTT GCA AAC GGC CAA GTC CAG ATC TTC AGC GGA AGA GGC
 Arg Gly Leu Arg Val Ala Asn Gly Gln Val Gln Ile Phe Ser Gly Arg Gly

3080 3090 3100 3110 3120
 ACC GCC ATC GGC ACT GAT AGC AGC CTC ACT CTC AAC ATC CGG GCG CCC CTA
 Thr Ala Ile Gly Thr Asp Ser Ser Leu Thr Leu Asn Ile Arg Ala Pro Leu

FIG. 7K

31/51

3130 3140 3150 3160 3170
 CAA TTT TCT GGA CCC GCC TTG ACT GCT AGT TTG CAA GGC AGT GGT CCG ATT
 Gln Phe Ser Gly Pro Ala Leu Thr Ala Ser Leu Gln Gly Ser Gly Pro Ile

3180 3190 3200 3210 3220
 ACT TAC AAC AGC AAC AAT GGC ACT TTC GGT CTC TCT ATA GGC CCC GGA ATG
 Thr Tyr Asn Ser Asn Asn Gly Thr Phe Gly Leu Ser Ile Gly Pro Gly Met

3230 3240 3250 3260 3270
 TGG GTA GAC CAA AAC AGA CTT CAG GTA AAC CCA GGC GCT GGT TTA GTC TTC
 Trp Val Asp Gln Asn Arg Leu Gln Val Asn Pro Gly Ala Gly Leu Val Phe

3280 3290 3300 3310 3320
 CAA GGA AAC AAC CTT GTC CCA AAC CTT GCG GAT CCG CTG GCT ATT TCC GAC
 Gln Gly Asn Asn Leu Val Pro Asn Leu Ala Asp Pro Leu Ala Ile Ser Asp

3330 3340 3350 3360 3370
 AGC AAA ATT AGT CTC AGT CTC GGT CCC GGC CTG ACC CAA GCT TCC AAC GCC
 Ser Lys Ile Ser Leu Ser Leu Gly Pro Gly Leu Thr Gln Ala Ser Asn Ala

3380 3390 3400 3410 3420
 CTG ACT TTA AGT TTA GGA AAC GGG CTT GAA TTC TCC AAT CAA GCC GTT GCT
 Leu Thr Leu Ser Leu Gly Asn Gly Leu Glu Phe Ser Asn Gln Ala Val Ala

FIG. 7L

32/51

3430 3440 3450 3460 3470 3480
 ATA AAA GCG GGC CGG GGC TTA CGC TTT GAG TCT TCC TCA CAA GCT TTA GAG
 Ile Lys Ala Gly Arg Gly Leu Arg Phe Glu Ser Ser Ser Gln Ala Leu Glu

 3490 3500 3510 3520 3530
 AGC AGC CTC ACA GTC GGA AAT GGC TTA ACG CTT ACC GAT ACT GTG ATC CGC
 Ser Ser Leu Thr Val Gly Asn Gly Leu Thr Leu Thr Asp Thr Val Ile Arg

 3540 3550 3560 3570 3580
 CCC AAC CTA GGG GAC GGC CTA GAG GTC AGA GAC AAT AAA ATC ATT GTT AAG
 Pro Asn Leu Gly Asp Gly Leu Glu Val Arg Asp Asn Lys Ile Ile Val Lys

 3590 3600 3610 3620 3630
 CTG GGC GCG AAT CTT CGT TTT GAA AAC GGA GCC GTA ACC GCC ACC GTT
 Leu Gly Ala Asn Leu Arg Phe Glu Asn Gly Ala Val Thr Ala Gly Thr Val

 3640 3650 3660 3670 3680
 AAC CCT TCT GCG CCC GAG GCA CCA CCA ACT CTC ACT GCA GAA CCA CCC CTC
Asn Pro Ser Ala Pro Glu Ala Pro Pro Thr Leu Thr Ala Glu Pro Pro Leu

 3690 3700 3710 3720 3730
 CGA GCC TCC AAC TCC CAT CTT CAA CTG TCC CTA TCG GAG GGC TTG GTT GTG
 Arg Ala Ser Asn Ser His Leu Gln Leu Ser Leu Ser Glu Gly Leu Val Val

FIG. 7M

33/51

```

3740      3750      3760      3770      3780
CAT AAC AAC GCC CTT GCT CTC CAA CTG GGA GAC GGC ATG GAA GTA AAT CAG
His Asn Asn Ala Leu Ala Leu Gln Leu Gly Asp Gly Met Glu Val Asn Gln

3790      3800      3810      3820      3830
CAC GGA CTT ACT TTA AGA GTA GGC TCG GGT TTG CAA ATG CGT GAC GGC ATT
His Gly Leu Thr Leu Arg Val Gly Ser Gly Leu Gln Met Arg Asp Gly Ile

3840      3850      3860      3870      3880
TTA ACA GTT ACA CCC AGC GGC ACT CCT ATT GAG CCC AGA CTG ACT GCC CCA
Leu Thr Val Thr Pro Ser Gly Thr Pro Ile Glu Pro Arg Leu Thr Ala Pro

3890      3900      3910      3920      3930
CTG ACT CAG ACA GAG AAT GGA ATC GGG CTC GCT CTC GGC GCC GGC TTG GAA
Leu Thr Gln Thr Glu Asn Gly Ile Gly Leu Ala Leu Gly Ala Gly Leu Glu

3940      3950      3960      3970      3980      3990
TTA GAC GAG AGC GCG CTC CAA GTA AAA GTT GGG CCC GGC ATG CGC CTG AAC
Leu Asp Glu Ser Ala Leu Gln Val Lys Val Gly Pro Gly Met Arg Leu Asn

4000      4010      4020      4030      4040
CCT GTA GAA AAG TAT GTA ACC CTG CTC CTG GGT CCT GGC CTT AGT TTT GGG
Pro Val Glu Lys Tyr Val Thr Leu Leu Leu Gly Pro Gly Leu Ser Phe Gly

```

FIG. 7N

34/51

4050 CAG CCG GCC AAC AGG ACA AAT TAT GAT GTG CGC GTT TCT GTG GAG CCC CCC 4090
 Gln Pro Ala Asn Arg Thr Asn Tyr Asp Val Arg Val Ser Val Glu Pro Pro

 4100 ATG GTT TTC GGA CAG CGT GGT CAG CTC ACA TTT TTA GTG GGT CAC GGA CTA 4140
 Met Val Phe Gly Gln Arg Gly Gln Leu Thr Phe Leu Val Gly His Gly Leu

 4150 CAC ATT CAA AAT TCC AAA CTT CAG CTC AAT TTG GGA CAA GGC CTC AGA ACT 4190
 His Ile Gln Asn Ser Lys Leu Gln Leu Asn Leu Gly Gln Gly Leu Arg Thr

 4200 GAC CCC GTC ACC AAC CAG CTG GAA GTG CCC CTC GGT CAA GGT TTG GAA ATT 4240
 Asp Pro Val Thr Asn Gln Leu Glu Val Pro Leu Gly Gln Gly Leu Glu Ile

 4250 GCA GAC GAA TCC CAG GTT AGG GTT AAA TTG GGC GAT GGC CTG CAG TTT GAT 4290
 Ala Asp Glu Ser Gln Val Arg Val Lys Leu Gly Asp Gly Leu Gln Phe Asp

 4300 TCA CAA GCT CGC ATC ACT ACC GCT CCT AAC ATG GTC ACT GAA ACT CTG TGG 4340
 Ser Gln Ala Arg Ile Thr Thr Ala Pro Asn Met Val Thr Glu Thr Leu Trp

FIG. 70

35/51

4350 4360 4370 4380 4390
 ACC GGA ACA GGC AGT AAT GCT AAT GTT ACA TGG CGG GGC TAC ACT GCC CCC
 Thr Gly Thr Gly Ser Asn Ala Asn Val Thr Trp Arg Gly Tyr Thr Ala Pro

4400 4410 4420 4430 4440
 GGC AGC AAA CTC TTT TTG AGT CTC ACT CGG TTC AGC ACT GGT CTA GTT TTA
 Gly Ser Lys Leu Phe Leu Ser Leu Thr Arg Phe Ser Thr Gly Leu Val Leu

4450 4460 4470 4480 4490 4500
 GGA AAC ATG ACT ATT GAC AGC AAT GCA TCC TTT GGG CAA TAC ATT AAC GCG
 Gly Asn Met Thr Ile Asp Ser Asn Ala Ser Phe Gly Gln Tyr Ile Asn Ala

4510 4520 4530 4540 4550
 GGA CAC GAA CAG ATC GAA TGC TTT ATA TTG TTG GAC AAT CAG GGT AAC CTA
 Gly His Glu Gln Ile Glu Cys Phe Ile Leu Leu Asp Asn Gln Gly Asn Leu

4560 4570 4580 4590 4600
 AAA GAA GGA TCT AAC TTG CAA GGC ACT TGG GAA GTG AAG AAC AAC CCC TCT
 Lys Glu Gly Ser Asn Leu Gln Gly Thr Trp Glu Val Lys Asn Asn Pro Ser

4610 4620 4630 4640 4650
 GCT TCC AAA GCT GCT TTT TTG CCT TCC ACC GCC CTA TAC CCC ATC CTC AAC
 Ala Ser Lys Ala Ala Phe Leu Pro Ser Thr Ala Leu Tyr Pro Ile Leu Asn

FIG. 7P

36/51

4660 4670 4680 4690 4700
 GAA AGC CGA GGG AGT CTT CCT GGA AAA AAT CTT GTG GGC ATG CAA GCC ATA
Glu Ser Arg Gly Ser Leu Pro Gly Lys Asn Leu Val Gly Met Gln Ala Ile

 4710 4720 4730 4740 4750
 CTG GGA GGC GGC ACT TGC ACT GTG ATA GCC ACC CTC AAT GGC AGA CGC
 Leu Gly Gly Gly Thr Cys Thr Val Ile Ala Thr Leu Asn Gly Arg Arg

 4760 4770 4780 4790 4800
 AGC AAC AAC TAT CCC GCG GGC CAG TCC ATA ATT TTC GTG TGG CAA GAA TTC
 Ser Asn Asn Tyr Pro Ala Gly Gln Ser Ile Ile Phe Val Trp Gln Glu Phe

 4810 4820 4830 4840 4850
 AAC ACC ATA GCC CGC CAA CCT CTG AAC CAC TCT ACA CTT ACT TTT TCT TAC
 Asn Thr Ile Ala Arg Gln Pro Leu Asn His Ser Thr Leu Thr Phe Ser Tyr

 4860 4870 4880 4890 4900
 TGG ACT TA AAT AAG TTG GAA ATA AAG AGT TAA ACT GAA TGT TTA AGT GCA
 Trp Thr

 4910 4920 4930 4940 4950
 ACA GAC TTT TAT TGG TTT TGG CTC ACA ACA AAT TAC AAC AGC ATA GAC AAG

 4960 4970 4980 4990 5000
 TCA TAC CGG TCA AAC AAC ACA GGC TCT CGA AAA CGG GCT AAC CGC TCC AAG

FIG. 7Q

37/51

5010
AAT CTG TCA CGC AGA CGA GCA AGT CCT AAA TGT TTT TTC ACT CTC TTC GGG 5060

5020 5030 5040 5050
AAT CTG TCA CGC AGA CGA GCA AGT CCT AAA TGT TTT TTC ACT CTC TTC GGG

5070 5080 5090 5100
GCC AAG TTC AGC ATG TAT CGG ATT TTC TGC TTA CAC CTT T

FIG. 7R

38/51

Ad2	MSKEIPTPYMWSYQPQMGLAAGAAQDYSTRINYMSAGPHMISRVNGIRAH	50
BAV3	LIKQPVVGTTHV-----EMPRNEVLEQH	23
Ad2	RNRILLEQAAITTTTPRNNLNPRSWPAALVYQESPAPTTVVLPDQAQAEVQ	100
BAV3	LTSHGAQIAGGG-----AAGDYFKSPTSARTLIPLTASCL-----RPDG	62
Ad2	MTNSGAQLAGGFRHRVRSPPQGITHLKIRGRGIQLNDESVSSSLGLRPDG	150
BAV3	VFQLGGGSRSSFNPLQTDFAFHALPSRPRHGGIGSRQFVEEFVPAVYLN	112
Ad2	TFQIGGAGRSSFTPRQAILTLQTSSEPRSGGIGTLQFIEEFVPSVYFNP	200
BAV3	YSGPPDSYPDQFIRHYNVYSNSVSGYS	139
Ad2	FSGPPGHYPDQFIPNFDAVKDSADGYD	227

FIG. 8A

BAV3	M-----EPOGVHAEQQFILNQISCANTALQRQREELASLVMLHACKRGL	77
Ad5	MTDTLDLEMDGIITEQRLL--ERRRAAAEQQRMNQELQDMVNLHQCKRGI	48
BAV3	FCPVKTYKLSLNASASEHSLHFESPSRFTLVNTHAGASVRVALHHQGAS	127
Ad5	FCLVKQAKVTYDSNTTGHRLSYKLPTKRQKLVMVGEKPIITITQHSVETE	98
BAV3	GSIRCSCSHAECPLVLLKTLCAFNFLD	154
Ad5	GCIHSPCQGPEDLCTLIKTLGLKDLIPFN	128

FIG. 8B

40/51

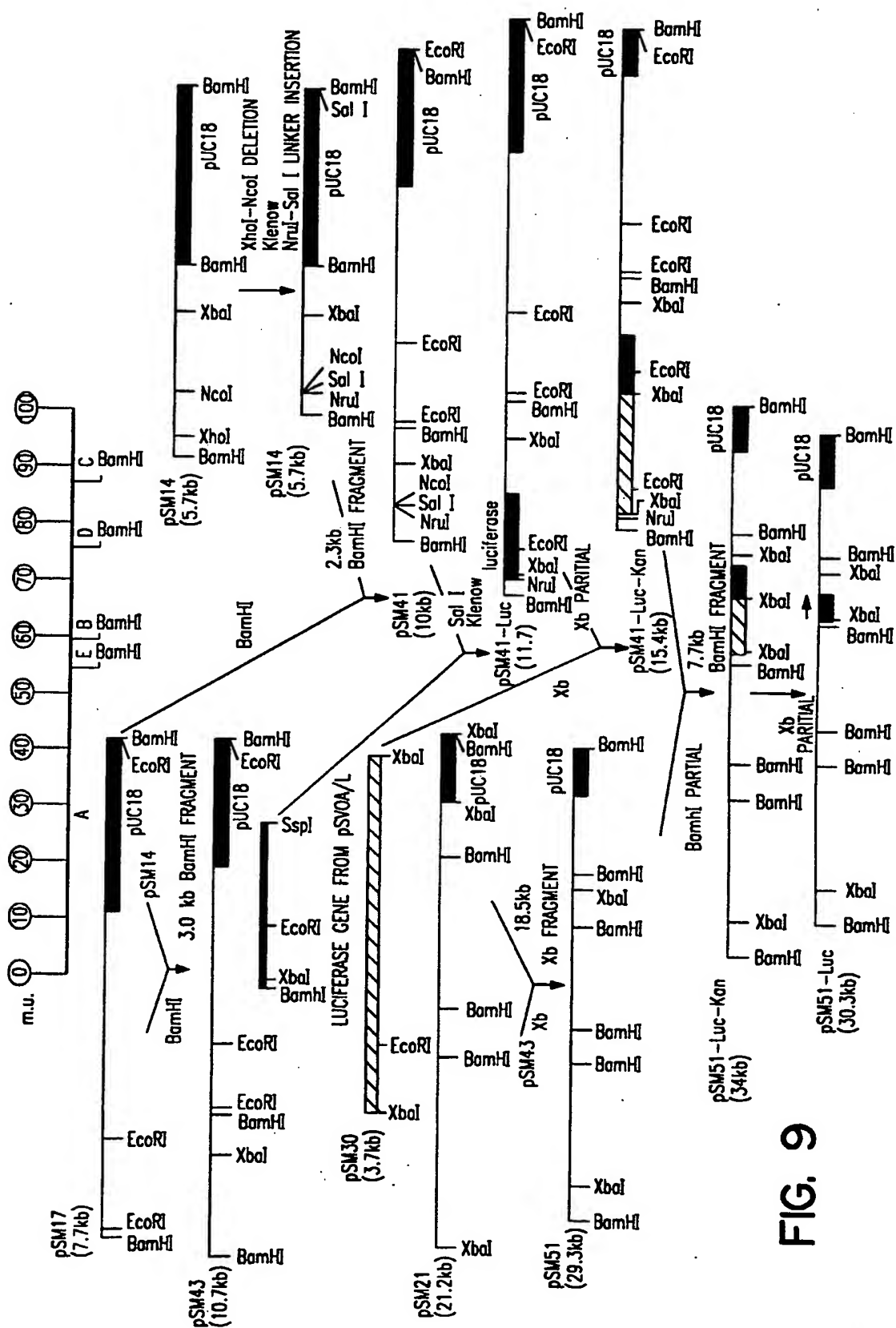
BAV3 - AKLGPGGLGTDDSGRSVVRTGRGLRVANGQVQIFSGRGTAIGTDSSLTLNI -392
 Ad2 - TKVAGAIGYDSSNNMEIKTGGGMRINNLL--LILDVDYPFDAQTKLRRLKL -284
 BAV3 - RAPLQFSGPALTASLQSGSPITYNSNNGTFGLSIGPGMWVDQNRLQVNP -442
 Ad2 - -----GQGPLYINASHN-----LDINYN -302
 BAV3 - AGLVFQGNLVPNLADPLAISDSKISLSLGPGLTQASNALTSLGNGLEF -492
 Ad2 - RGLYL-----FNASNNTKKLEVSIIKKSS-----GLNF -329
 BAV3 - SNQAVAIKAGRGLRFESSQALESSTVGNGLTLTDTVIRPNLGDGLEVR -542
 Ad2 - DNTAIAINAGKGLEFDTNT----- -348
 BAV3 - DNKIIIVKLGANLRFENGAVTAGTVNPSAPEAPPTLTAEPPLRASNHLQL -592
 Ad2 - ----- -348
 BAV3 - SLSEGLVHNNALALQLGDGMEVNQHGLTLRVGSGLQMRDGILTVTPSGT -642
 Ad2 - -----SESPDIN--PIKTKIGSGID-----YNENGA -372
 BAV3 - PIEPRLTAPLTQTENGIGLALGAGLELDESALQVKVGPGRMLNPVEKYVT -692
 Ad2 - MIT-----KLGAGLSFDNSG----- -387

FIG. 8C-2

BAV3 - LLLGPGLSFGQPANRTNYDVRVSVEPPMVFGQRGQLTFLVGHGLHIQNSK -742
 Ad2 - -----AITIG-----NKNDKLTLTWTTDPDPS-----NCR -412
 BAV3 - LQLNLGQGLRTDPVTNQLEVP LGQGLEIADESQVRVKLGDGLQFDSQARI -792
 Ad2 - IHSD-----NDCKFTLVLT---KCGSQVLA -434
 BAV3 - TTAPNMVTETLWTGTGSNANVTWRGYTAPGSKLFLSLTRFSTGLVLGNMT -842
 Ad2 - TVAALAVSGDLSSMTGTVASVS-----IFLRFDQ--NGVLMENSS -472
 BAV3 - IDSNASFGQYINAGHEQIECFILLDNQGNLKEGSNLQGTWEVKNNPSASK -892
 Ad2 - LKKHY-----WNFRNGNS-----TNANPYTNA -494
 BAV3 - AAFLPSTALYPILNESRGSPLPGKNLVGMQAILGGGGTCTVIA--TLNGRRS -941
 Ad2 - VGFMNLLAYP---KTQSQTAKNNIVSQVYLHGDKTKPMILTITLNGTSE -541
 BAV3 - NNYPAGQSII---FVWQ--EFNTIARQPLNHSTLTFSYWT -976
 Ad2 - STETSEVSTYSMSFTWSWESGKYTTETFATNSYTF SYIAQE -582

FIG. 8C-3

SUBSTITUTE SHEET (RULE 26)



১৫৬

42/51

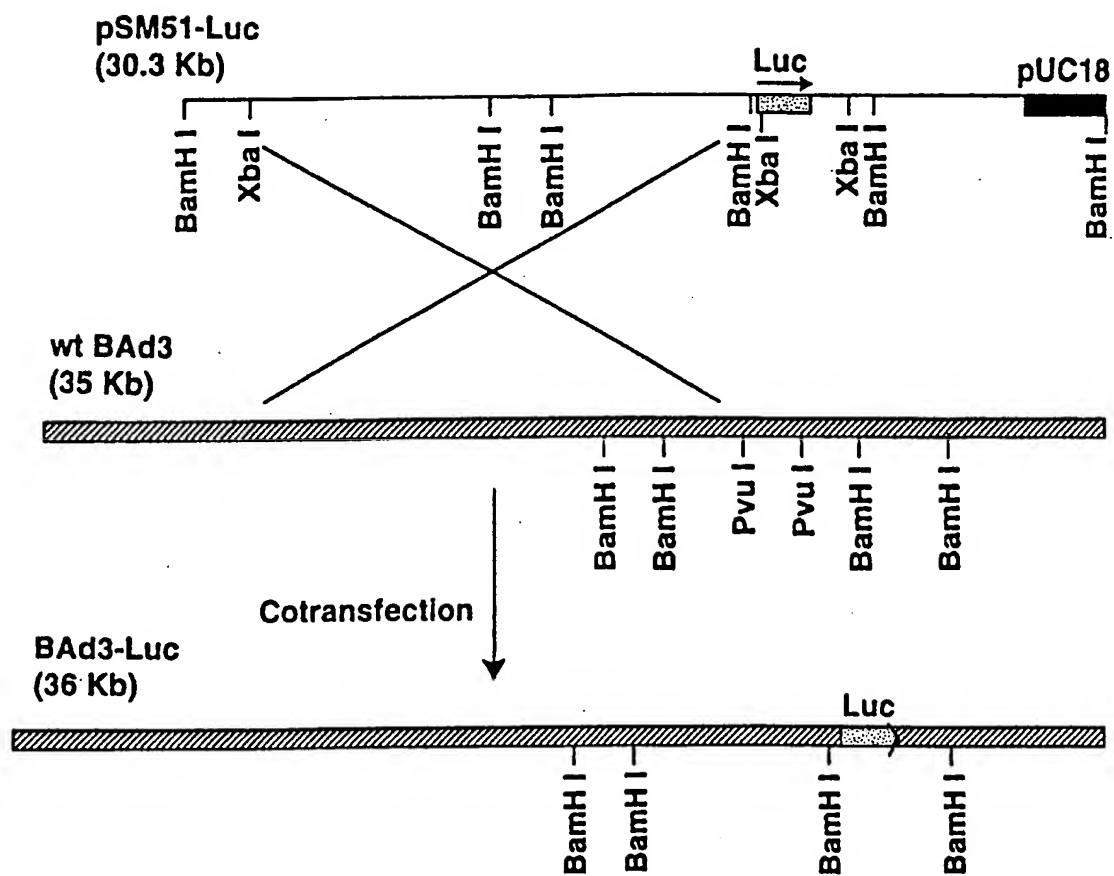


FIG. 10

Mock				3.1				3.2				wt			
B	E	Xb	B	E	Xb	B	E	Xb	B	E	Xb	B	E	Xb	B

18.5 ←
8.1 ←
6.0 →
4.0 →
3.2 →
2.9 →

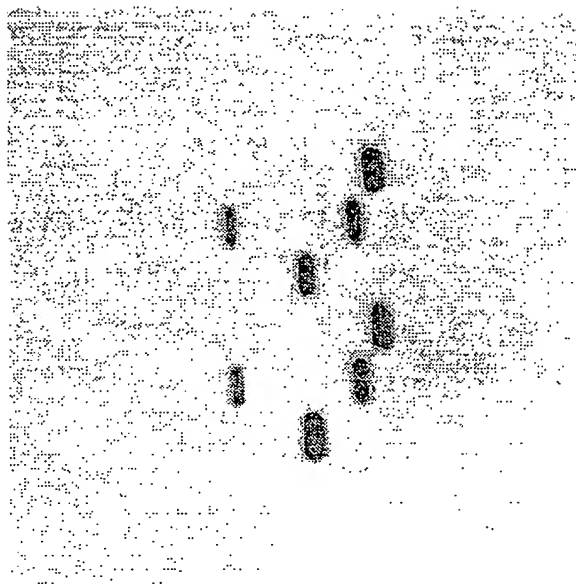


FIG. IIB

Mock				3.1				3.2				wt			
B	E	Xb	B	E	Xb	B	E	Xb	B	E	Xb	B	E	Xb	B

18.5 ←
8.1 ←
6.0 →
4.0 →
3.2 →
2.9 →

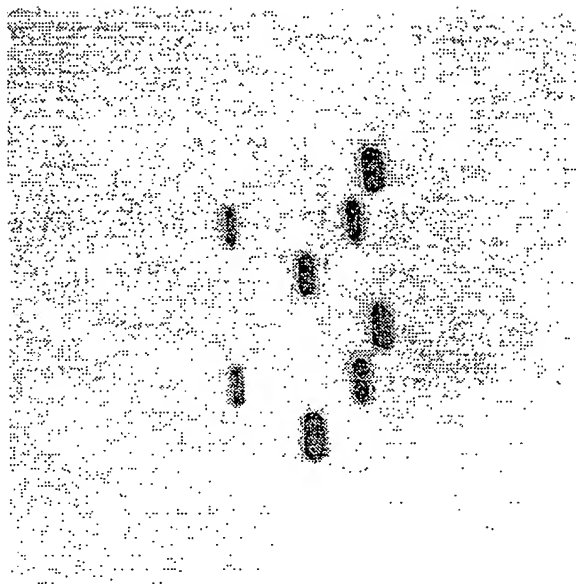


FIG. IIA

44/51

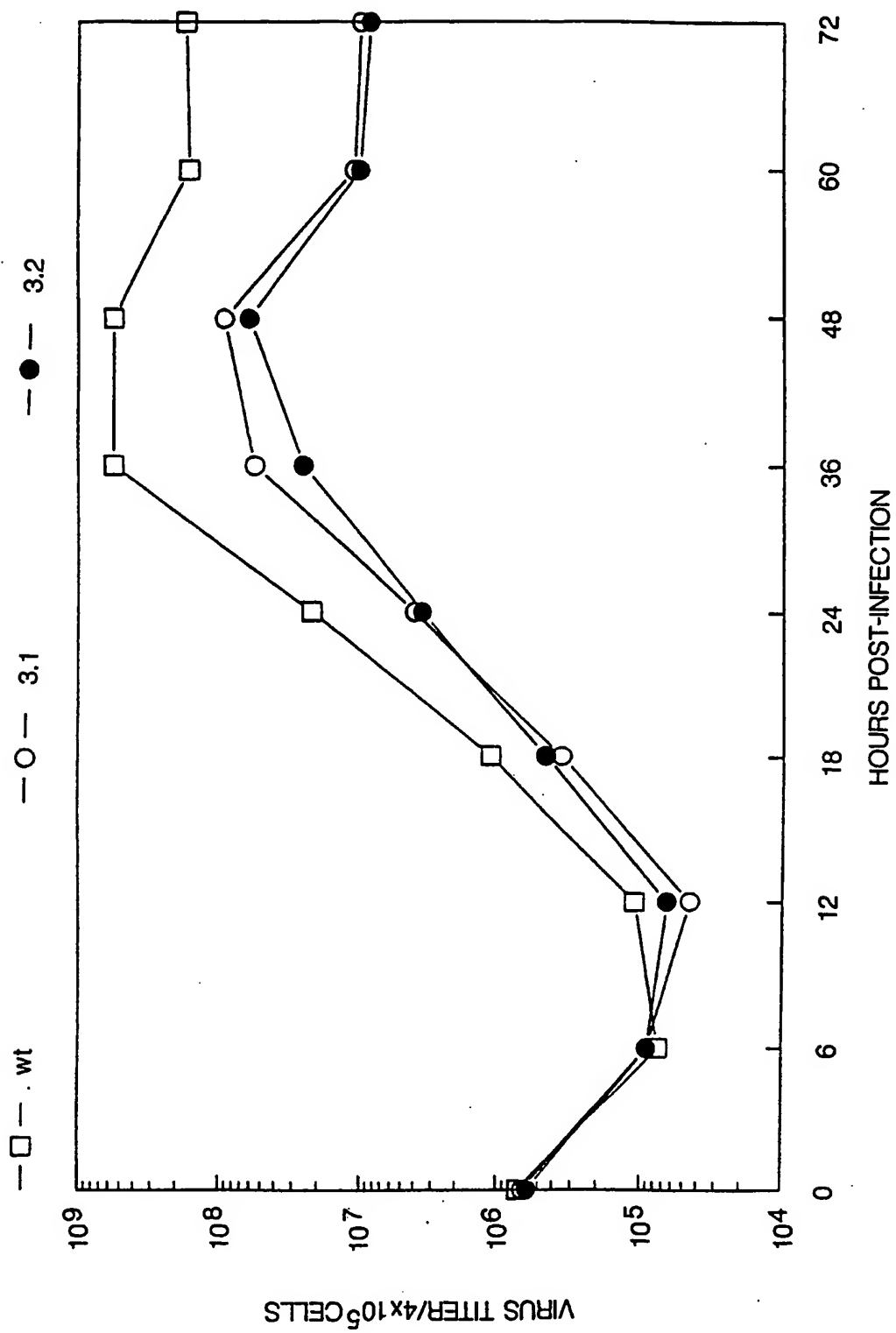


FIG. 12

45/51

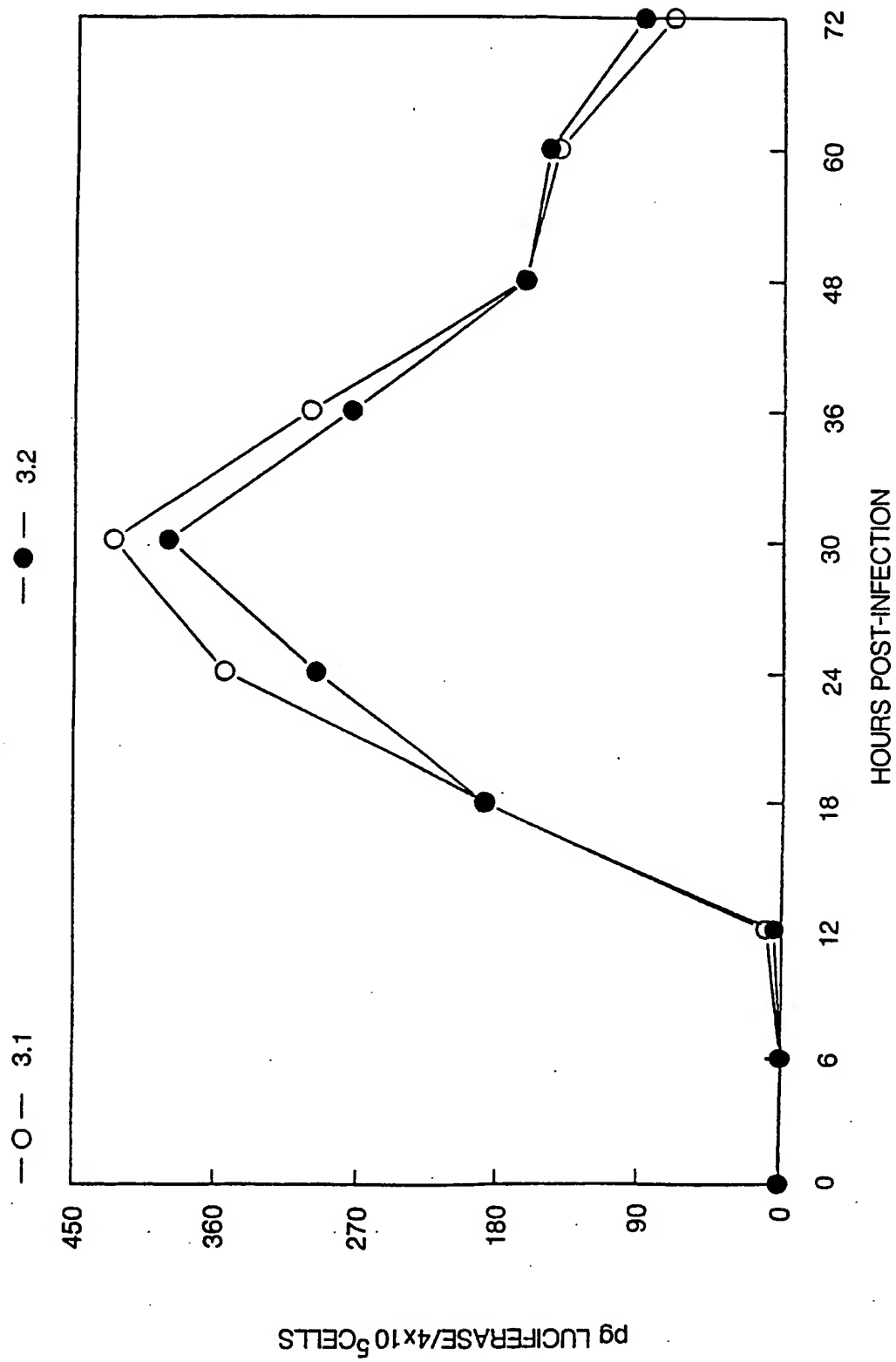


FIG. 13

pg LUCIFERASE/4x10⁵ CELLS

SUBSTITUTE SHEET (RULE 26)

46/51

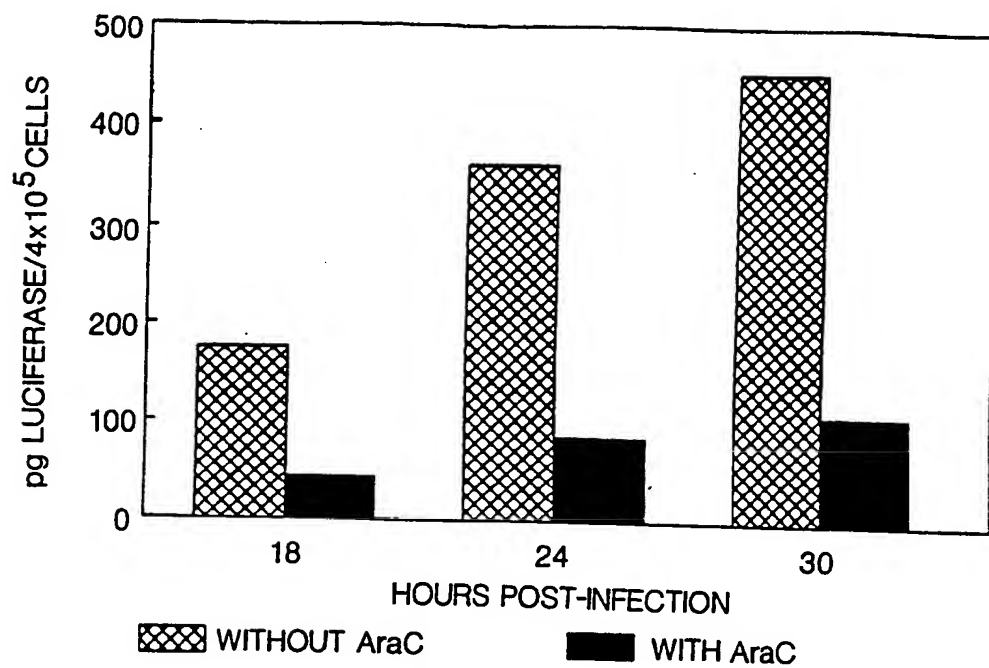


FIG. 14A

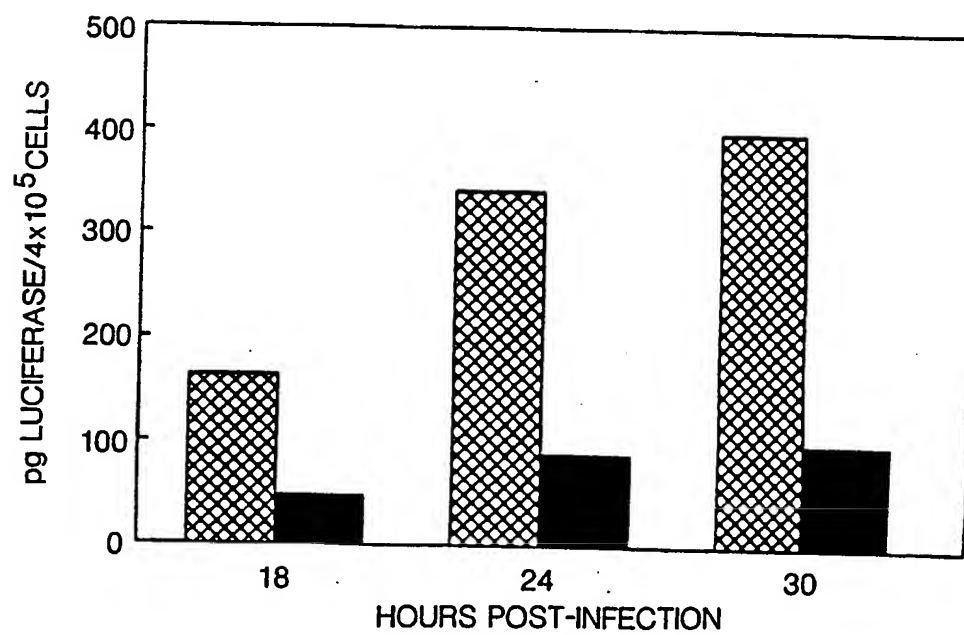


FIG. 14B

47/51

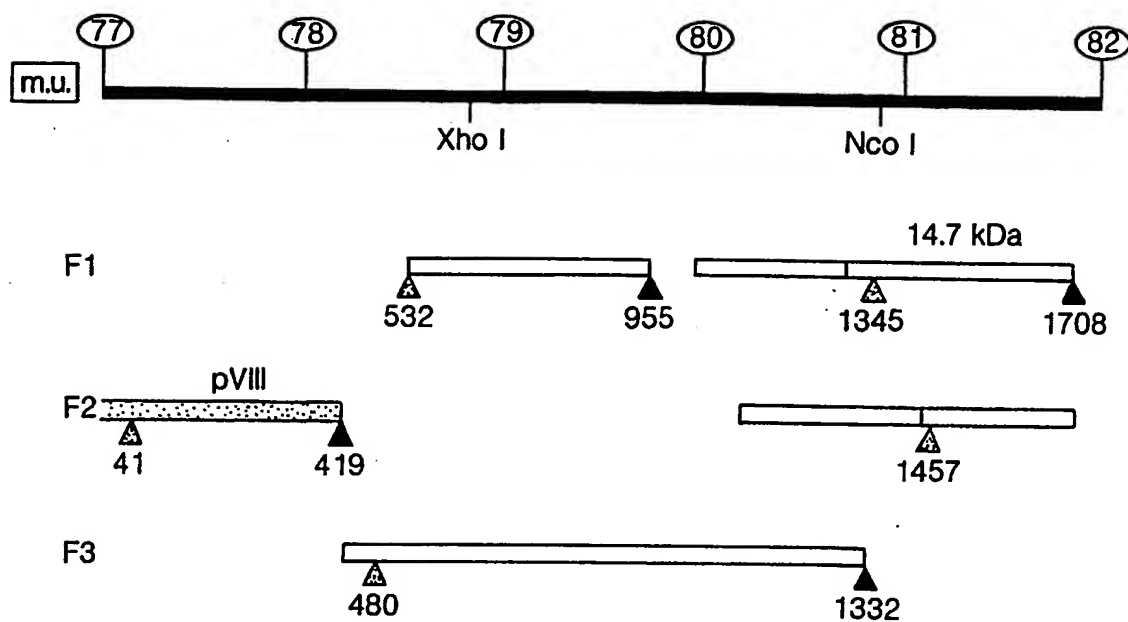


FIG. 15A

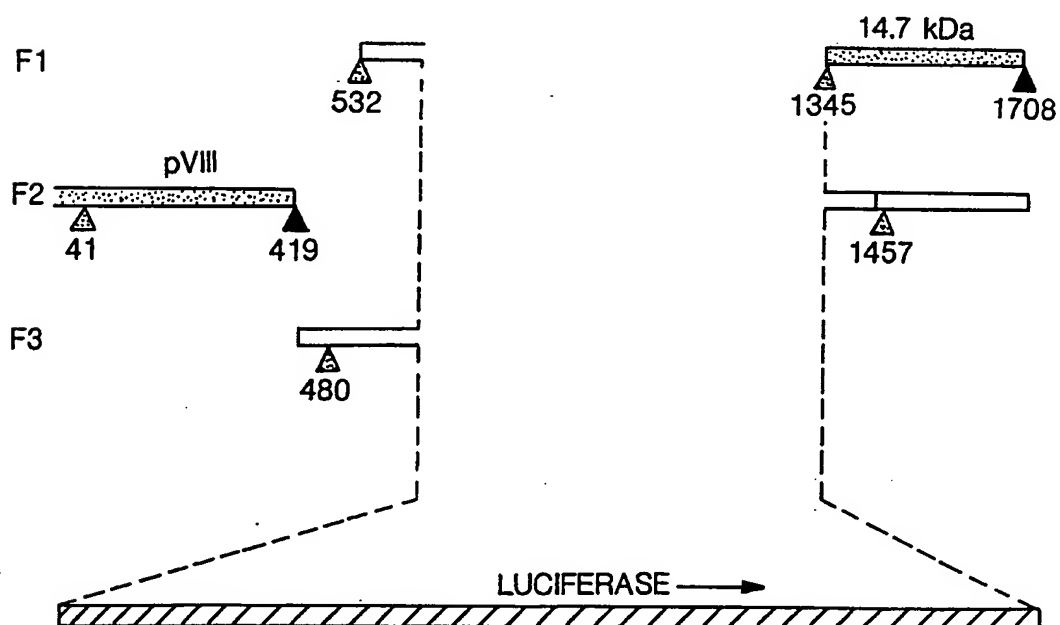


FIG. 15B

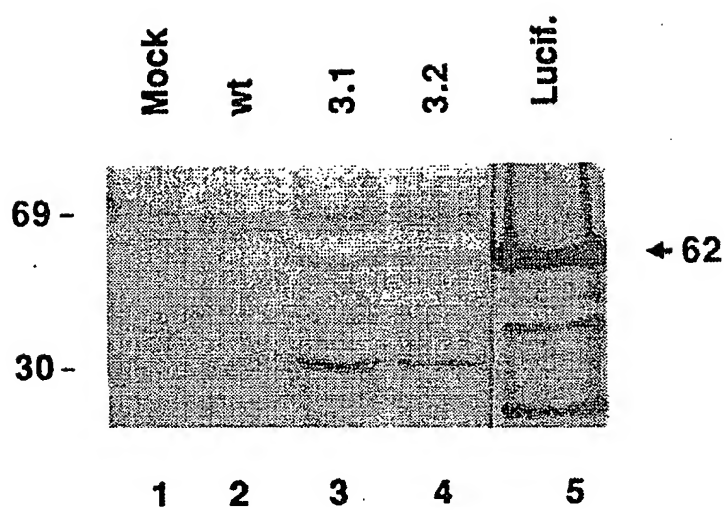


FIG. 16

49/51

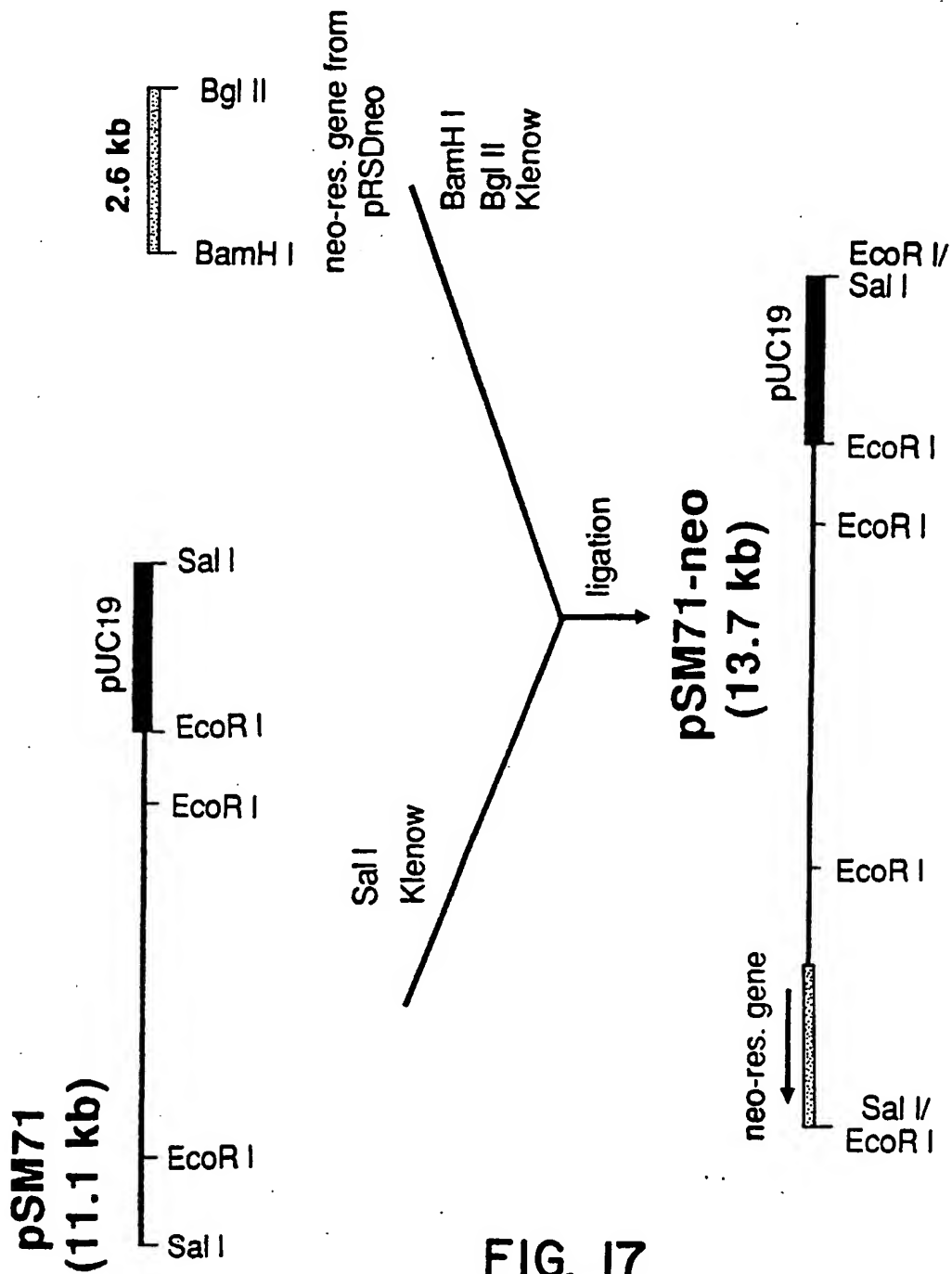


FIG. 17

50/51

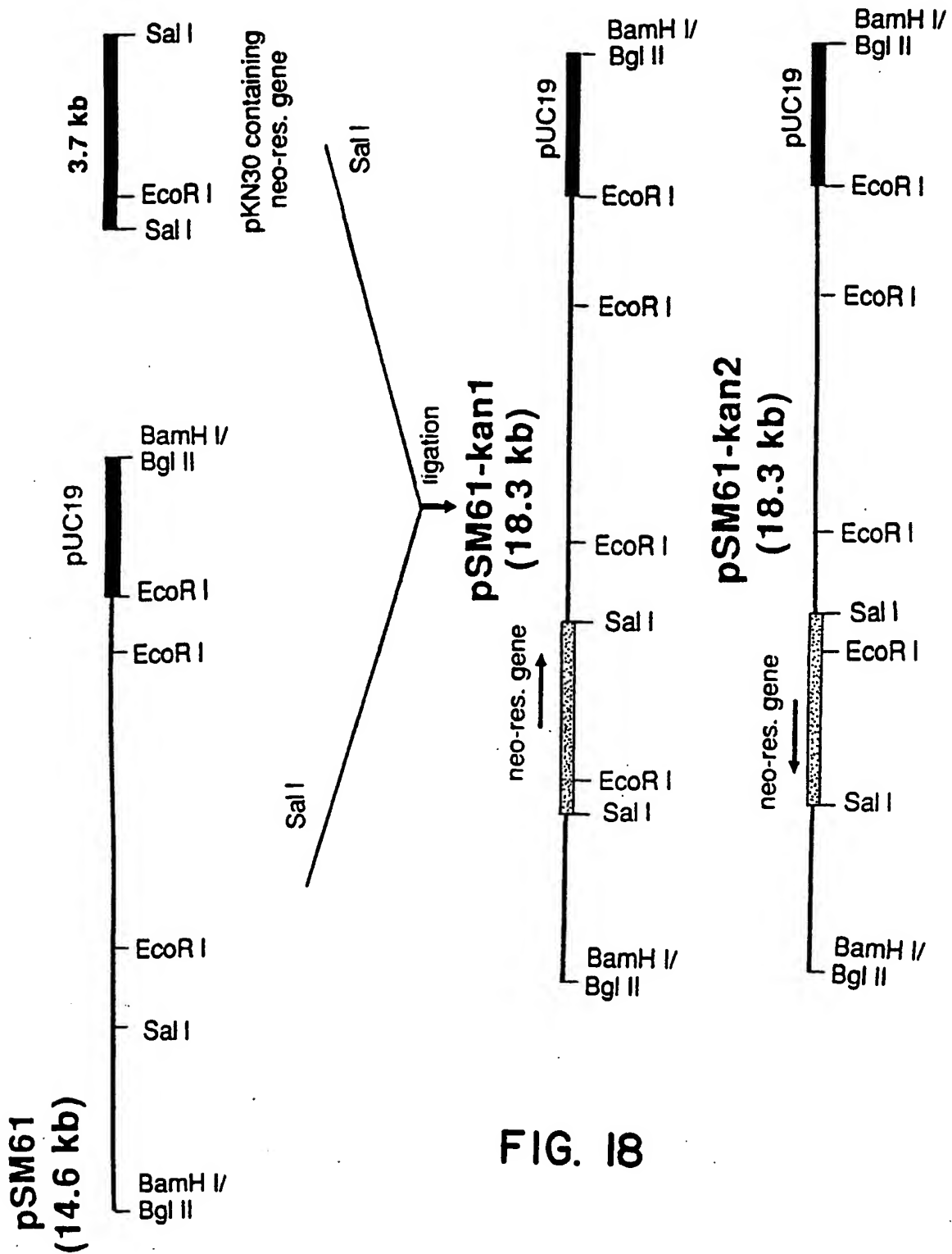


FIG. 18

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶: C12N 15/86, 5/10, A61K 48/00, C12R 1/92	A3	(11) International Publication Number: WO 95/16048 (43) International Publication Date: 15 June 1995 (15.06.95)
(21) International Application Number: PCT/CA94/00678 (22) International Filing Date: 9 December 1994 (09.12.94) (30) Priority Data: 08/164,292 9 December 1993 (09.12.93) US (71) Applicant: UNIVERSITY OF SASKATCHEWAN [CA/CA]; 124 Veterinary Road, Saskatoon, Saskatchewan S5R 2Y4 (CA). (72) Inventors: MITTAL, Suresh, K.; #201-235 Kingsmere Boulevard, Saskatoon, Saskatchewan S7J 4J6 (CA). GRAHAM, Frank, L.; 34 Amelia Street, Hamilton, Ontario L8P 2V4 (CA). PREVEC, Ludvik; 944 LaSalle Park Road, Burlington, Ontario L7T 1M9 (CA). BABIUK, Lorne, A.; 245 East Place, Saskatoon, Saskatchewan S7J 2Y1 (CA). (74) Agent: VAN ZANT, Joan, M.; Scott & Ayles, 60 Queen Street, Ottawa, Ontario K1P 5Y7 (CA).		(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i> (88) Date of publication of the international search report: 28 September 1995 (28.09.95)
(54) Title: RECOMBINANT PROTEIN PRODUCTION IN BOVINE ADENOVIRUS EXPRESSION VECTOR SYSTEM		
(57) Abstract		
<p>The present invention relates to novel live bovine adenovirus (BAV) expression vector systems in which part or all of one or both of the early region 1 (E1) and early region 3 (E3) genes are deleted and replaced by a foreign gene or fragment thereof and novel recombinant mammalian cell lines stably transformed with BAV E1 sequences, and therefore, express E1 gene products capable of allowing replication, therein of a bovine adenovirus having an E1 deletion replaced by a heterologous nucleotide sequence encoding a foreign gene or fragment thereof and their use in production of (antigenic) polypeptides or fragments thereof for the purpose of live recombinant virus or subunit vaccine or for other therapies.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Larvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

INTERNATIONAL SEARCH REPORT

International Application No

PCT/CA 94/00678

A. CLASSIFICATION OF SUBJECT MATTER
 IPC 6 C12N15/86 C12N5/10 A61K48/00 C12R1/92

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US,A,4 920 209 (DAVIES ET AL.) 24 April 1990 *see the whole patent* ---	1-23
Y	JOURNAL OF VIROLOGY, vol. 57, no. 1, 1986 pages 267-274, Y. HAJ-AHMAD ET AL. 'Development of a helper-independent human adenovirus vector and its use in the transfer of HSV-tk gene' *see the whole article* ---	1-23
Y	Journal of cellular Biochemistry, UCLA symposium on Molecular and Cellular Biology, A.R. Liss, Inc. New York, 1988, Suppl. 12B, F109 *see the whole abstract* ---	1-23
-/--		

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *&* document member of the same patent family

Date of the actual completion of the international search

12 July 1995

Date of mailing of the international search report

29.08.1995

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl,
 Fax: (+ 31-70) 340-3016

Authorized officer

Marie, A

INTERNATIONAL SEARCH REPORT

International Application No

PCT/CA 94/00678

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JOURNAL OF GENERAL VIROLOGY, vol. 70, 1989 pages 429-434, L. PREVEC ET AL. 'Use of human adenovirus-based vectors for antigen expression in animals' *see the whole article* ----	1-23
Y	JOURNAL OF GENERAL VIROLOGY, vol. 73, 1992 pages 3295-3300, S.K. MITTAL ET AL. 'Sequence analysis of bovine adenovirus type 3 early region 3 and fibre protein genes' *see the whole article* ----	1-23
Y	JOURNAL OF VIROLOGY, vol. 49, no. 2, 1984 pages 604-608, S.-L. HU ET AL. 'Sequence homology between bovine and human adenoviruses' *see the whole article* -----	1-23

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/CA 94/00678

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US-A-4920209	24-04-90	AU-B- 576907	08-09-88
		AU-A- 4884085	08-05-86
		CA-A- 1263305	28-11-89
		DE-A- 3586841	24-12-92
		EP-A, B 0181117	14-05-86
		GB-A, B 2166349	08-05-86
		IE-B- 58205	28-07-93
		JP-A- 61118326	05-06-86
		KR-B- 9312116	24-12-93
